STIC-Biotech/ChemLib 181150

From: Whiteman, Brian
Sent: Friday, March 03, 2006 9:19 AM
To: STIC-Biotech/ChemLib
Subject: seq search

10/782,899

nucleotides 1-192 of SEQ ID NO: 1 and nucleotides 472-600 of SEQ ID NO: 1
1) us patents and published us patent application databases

Thank you,

Brian Whiteman
Remsen, 2D14
mail box 2C18
Patent Examiner - Art Unit 1635
United States Patent and Trademark Office
(571) 272-0764

RECEIVED
MAR - 3 2006
[illegible]
(STIC)

****
Searcher:_______________
Searcher Phone:___________
Date Searcher Picked up:_____
Date completed:___________
Searcher Prep Time:_______
Online Time:_______________

****
Type of Search
NA#_______ AA#:_______
S/L:_____ Oligomer:______
Encode/Transl:___________
Structure #:______Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:________________
DIALOG:______________
QUESTEL/ORBIT:_________
LEXIS/NEXIS:___________
SEQUENCE SYSTEM:_______
WWW/Internet:__________
Other (Specify):__________

```
                        GenCore version 5.1.7
               Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 11:25:13 ; Search time 1781.23 Seconds
                                           (without alignments)
                                           6127.182 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:      11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_in:*
                3:  gb_env:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pr:*
                9:  gb_ro:*
                10:  gb_sts:*
                11:  gb_sy:*
                12:  gb_un:*
                13:  gb_vi:*
                14:  gb_htg:*
                15:  gb_pl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                 %
Result           Query
   No.   Score   Match Length DB  ID                   Description
 ----------------------------------------------------------------------------
     1     192   100.0    600  6  BD144660             BD144660 Drug for
 c   2   175.2    91.2  56640  1  AE014295_22          Continuation (23 o
 c   3   175.2    91.2 156638  6  AX492788             AX492788 Sequence
 c   4   175.2    91.2 156646  6  AX553955             AX553955 Sequence
     5   173.2    90.2    600  1  AB072446             AB072446 Bifidobac
 c   6      37    19.3 230625  14  AC096816             AC096816 Rattus no
 c   7      37    19.3 246991  14  AC122956             AC122956 Rattus no
     8    36.4    19.0  67395  8  AL671518             AL671518 Human DNA
     9    36.4    19.0 317403  14  AC154842             AC154842 Mus muscu
    10    35.8    18.6 170350  8  AC140952             AC140952 Pan trogl
    11    35.8    18.6 171239  14  AC092873             AC092873 Pan trogl
    12    35.8    18.6 171391  14  AC098646             AC098646 Pan trogl
 c  13    35.8    18.6 183888  8  AC151799             AC151799 Pan trogl
 c  14      35    18.2 181622  14  AC149460             AC149460 Papio anu
    15      35    18.2 231844  14  AC116068             AC116068 Rattus no
 c  16      35    18.2 234234  14  AC097194             AC097194 Rattus no
    17    34.8    18.1 152430  8  AC087432             AC087432 Homo sapi
    18    34.8    18.1 160410  14  AC020753             AC020753 Homo sapi
```

```
    19    34.8    18.1 166162  8  AC018814             AC018814 Homo sapi
 c  20    34.8    18.1 181870  14  AC024216             AC024216 Homo sapi
    21    34.8    18.1 188432  14  AC018497             AC018497 Homo sapi
    22    34.4    17.9 148269  14  AC026486             AC026486 Homo sapi
    23    34.4    17.9 162249  14  AC020682             AC020682 Homo sapi
 c  24    34.4    17.9 166973  8  AL451049             AL451049 Human DNA
 c  25    34.2    17.8 155711  2  VYIVD10              AL360354 Plasmodiu
    26    34.2    17.8 183089  9  AL663115             AL663115 Mouse DNA
    27    34.2    17.8 189862  14  AC157771             AC157771 Mus muscu
    28    34.2    17.8 213720  9  AC141477             AC141477 Mus muscu
 c  29    34.2    17.8 235639  9  AC158618             AC158618 Mus muscu
 c  30      34    17.7 166873  8  AC008554             AC008554 Homo sapi
    31      34    17.7 179393  8  AC010636             AC010636 Homo sapi
    32    33.6    17.5   4799  9  AF367244             AF367244 Mus muscu
    33    33.6    17.5  17073  2  AC084160             AC084160 Caenorhab
    34    33.6    17.5 179650  9  AC158525             AC158525 Mus muscu
 c  35    33.4    17.4 114784  8  HS1189B24            AL030996 Human DNA
    36    33.2    17.3    630  10  G54937               G54937 SHGC-100131
    37    33.2    17.3   4765  9  BC052387             BC052387 Mus muscu
    38    33.2    17.3   5011  9  AF140786             AF140786 Mus muscu
    39    33.2    17.3   5034  9  AF387811             AF387811 Mus muscu
    40    33.2    17.3  74436  14  AC026625             AC026625 Homo sapi
    41    33.2    17.3  87258  8  AF091512             AF091512 Homo sapi
    42    33.2    17.3 110000  15  CR382122_05          Continuation (6 of
    43    33.2    17.3 132910  14  AC014319             AC014319 Drosophil
    44    33.2    17.3 160421  8  AC116546             AC116546 Homo sapi
    45    33.2    17.3 168700  14  AC009115             AC009115 Homo sapi

                                  ALIGNMENTS

RESULT 1
BD144660
LOCUS       BD144660                 600 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Drug for gene therapy using anaerobic bacteria.
ACCESSION   BD144660
VERSION     BD144660.1  GI:27850418
KEYWORDS    JP 2002097144-A/1.
SOURCE      Bifidobacterium longum
  ORGANISM  Bifidobacterium longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Fujimori,M., Taniguchi,S., Amano,J., Yazawa,K., Kano,Y.,
            Nakamura,T. and Sasaki,T.
  TITLE     Drug for gene therapy using anaerobic bacteria
  JOURNAL   Patent: JP 2002097144-A 1 02-APR-2002;
            JUN AMANO,MINORU FUJIMORI
COMMENT     OS   Bifidobacterium longum
            PN   JP 2002097144-A/1
            PD   02-APR-2002
            PF   21-SEP-2001 JP 2001290187
            PI   MINORU FUJIMORI,SHUNICHIRO TANIGUCHI,JUN
            AMANO,KAZUTORA YAZAWA,
            PI   YASUMASA KANO,TOSHIYUKI NAKAMURA,TAKAYUKI SASAKI PC
            A61K35/74,A61K31/396,A61K31/7068,A61K31/711,A61K38/00,A61K38/ PC
            43,A61K48/00,
            PC   A61P35/00,C12N1/21,C12N15/09//(C12N1/21,C12R1:01),C12N15/00,
            PC   A61K37/02,
            PC   A61K37/48
            CC   Drug for gene therapy using anaerobic bacteria FH   Key
                 Location/Qualifiers
            FT   CDS                 (193)..(471).
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:216816"
ORIGIN

  Query Match             100.0%;  Score 192;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-49;
```

```
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 2
AE014295_22/c
WPCOMMENT
  Sequence split into 23 fragments  LOCUS AE014295 Accession AE014295
     Fragment Name      Begin      End
     AE014295_00            1    110000
     AE014295_01       100001    210000
     AE014295_02       200001    310000
     AE014295_03       300001    410000
     AE014295_04       400001    510000
     AE014295_05       500001    610000
     AE014295_06       600001    710000
     AE014295_07       700001    810000
     AE014295_08       800001    910000
     AE014295_09       900001   1010000
     AE014295_10      1000001   1110000
     AE014295_11      1100001   1210000
     AE014295_12      1200001   1310000
     AE014295_13      1300001   1410000
     AE014295_14      1400001   1510000
     AE014295_15      1500001   1610000
     AE014295_16      1600001   1710000
     AE014295_17      1700001   1810000
     AE014295_18      1800001   1910000
     AE014295_19      1900001   2010000
     AE014295_20      2000001   2110000
     AE014295_21      2100001   2210000
     AE014295_22      2200001   2256640
Continuation (23 of 23) of AE014295 from base 2200001 (AE014295 Bifidobacterium longum N

  Query Match             91.2%;  Score 175.2;  DB 1;  Length 56640;
  Best Local Similarity   97.4%;  Pred. No. 2.6e-44;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41859 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 41800

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41799 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 41740

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41739 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 41680

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db      41679 AAGAAGGATGCTTT 41666


RESULT 3
AX492788/c
LOCUS       AX492788               156638 bp    DNA     linear   PAT 26-SEP-2002
```

```
DEFINITION  Sequence 1106 from Patent EP1227152.
ACCESSION   AX492788
VERSION     AX492788.1  GI:23338471
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   None.
  TITLE     Bacterial strain and genome of bifidobacterium
  JOURNAL   Patent: EP 1227152-A 1106 31-JUL-2002;
            Societe des Produits Nestle S.A. (CH)
FEATURES             Location/Qualifiers
     source          1..156638
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="1106 seq in place of 1099 because seq 1 (2.256.638
                     bases splitted
                     in 7 more sequences.
                     seq 0001: from 0.000.001 to 0.349.980 length: 349980
                     seq 1100: from 0.300.001 to 0.649.980 length: 349980
                     seq 1101: from 0.600.001 to 0.949.980 length: 349980
                     seq 1102: from 0.900.001 to 1.249.980 length: 349980
                     seq 1103: from 1.200.001 to 1.549.980 length: 349980
                     seq 1104: from 1.500.001 to 1.849.980 length: 349980
                     seq 1105: from 1.800.001 to 2.149.980 length: 349980
                     seq 1106: from 2.100.001 to 2.256.638 length: 156638"
ORIGIN

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156638;
  Best Local Similarity   97.4%;  Pred. No. 2.6e-44;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142164 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142105

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142104 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142045

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142044 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141985

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db     141984 AAGAAGGATGCTTT 141971


RESULT 4
AX553955/c
LOCUS       AX553955               156646 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 9 from Patent WO02074798.
ACCESSION   AX553955
VERSION     AX553955.1  GI:25897908
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Arigoni,F., Delley,M., Mollet,B., Pridmore,R.D., Schell,M.A.,
            Pohl,T.G. and Zwahlen,M.C.
  TITLE     The genome of a bifidobacterium
  JOURNAL   Patent: WO 02074798-A 9 26-SEP-2002;
            SOCIETE DES PRODUITS NESTLE S.A. (CH)
FEATURES             Location/Qualifiers
     source          1..156646
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
```

```
                     /db_xref="taxon:1679"
                     /note="seq 1: original length 2256646 splitted into
                     seq 1: from 0.000.001 to 0.349.980
                     seq 3: from 0.300.001 to 0.649.980
                     seq 4: from 0.600.001 to 0.949.980
                     seq 5: from 0.900.001 to 1.249.980
                     seq 6: from 1.200.001 to 1.549.980
                     seq 7: from 1.500.001 to 1.849.980
                     seq 8: from 1.800.001 to 2.149.980
                     seq 9: from 2.100.001 to 2.256.646"
ORIGIN

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156646;
  Best Local Similarity   97.4%;  Pred. No. 2.6e-44;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142172 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142113

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142112 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142053

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142052 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141993

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db     141992 AAGAAGGATGCTTT 141979


RESULT 5
AB072446
LOCUS       AB072446                 600 bp    DNA     linear   BCT 15-JAN-2003
DEFINITION  Bifidobacterium longum biovar Longum hup gene for histone-like HU
            protein HB1, complete cds.
ACCESSION   AB072446
VERSION     AB072446.1  GI:19918953
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Takeuchi,A., Matsumura,H. and Kano,Y.
  TITLE     Cloning and expression in Escherichia coli of a gene, hup, encoding
            the histone-like protein HU of Bifidobacterium longum
  JOURNAL   Biosci. Biotechnol. Biochem. 66 (3), 598-603 (2002)
   PUBMED   12005055
REFERENCE   2  (bases 1 to 600)
  AUTHORS   Kano,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) Yasunobu Kano, Kyoto Pharmaceutical
            University, Institute of Molecular and Cellular Biology for
            Pharmaceutical Sciences; 1, Shichono-cho, Misasagi, Yamashina-ku,
            Kyoto 6078412, Japan (E-mail:ykano@mb.kyoto-phu.ac.jp,
            Tel:81755954719, Fax:81755954799)
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1679"
     RBS             180. .185
     gene            193. .474
                     /gene="hup"
     CDS             193. .474
                     /gene="hup"
                     /codon_start=1
                     /transl_table=11
                     /product="histone-like HU protein HB1"
                     /protein_id="BAB88690.1"
```

```
                     /db_xref="GI:19918954"
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRPARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
ORIGIN

  Query Match             90.2%;  Score 173.2;  DB 1;  Length 600;
  Best Local Similarity   97.4%;  Pred. No. 1.3e-43;
  Matches  187;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          3 TGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTC 60

Qy         63 --AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
                |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 6
AC096816/c
LOCUS       AC096816               230625 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-171E18, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AC096816
VERSION     AC096816.6  GI:30520628
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 230625)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
```

```
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230625)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 230625)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942762.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFJV
                Center clone name: CH230-171E18
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 197146 bases at least Q40
                Consensus quality: 202146 bases at least Q30
                Consensus quality: 205589 bases at least Q20
                Estimated insert size: 207699; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6876: contig of 6876 bp in length
            *     6877     6976: gap of unknown length
            *     6977    35044: contig of 28068 bp in length
            *    35045    35144: gap of unknown length
            *    35145    60636: contig of 25492 bp in length
            *    60637    60736: gap of unknown length
            *    60737    72181: contig of 11445 bp in length
            *    72182    72281: gap of unknown length
            *    72282   227314: contig of 155033 bp in length
            *   227315   227414: gap of unknown length
            *   227415   228683: contig of 1269 bp in length
            *   228684   228783: gap of unknown length
            *   228784   230625: contig of 1842 bp in length.
FEATURES             Location/Qualifiers
     source          1..230625
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-171E18"
     gap             6877..6976
                     /estimated_length=unknown
     gap             35045..35144
                     /estimated_length=unknown
     gap             60637..60736
                     /estimated_length=unknown
     gap             72182..72281
                     /estimated_length=unknown
     misc_feature    224079..227314
                     /note="wgs_contig"
     gap             227315..227414
                     /estimated_length=unknown
     gap             228684..228783
                     /estimated_length=unknown
ORIGIN

  Query Match             19.3%;  Score 37;  DB 14;  Length 230625;
  Best Local Similarity   54.9%;  Pred. No. 1.5;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          29 GACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAA 88
               || || |  | |  ||||| |    |||||    |   |   |    |||   |  | ||
Db      127530 GAAAAACCCATTTATGTCTAAGAGCTCTATGCAGATATCTAACTTCTAAAGGAACATAAA 127471

Qy          89 AACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
               ||  ||| | ||||  |  |||||  || |  ||| | ||||||  ||  | |   |  |
Db      127470 AATACTTTTCAAACATGTATTTTCTTAGTAGGATGAGGTAGTATACTTCTTTAACCCTAG 127411

Qy         149 GACTAAGCAAAAG 161
               ||||| ||| | |
Db      127410 GACTAGGCAGAGG 127398


RESULT 7
AC122956/c
LOCUS       AC122956              246991 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-17E8, WORKING DRAFT SEQUENCE.
ACCESSION   AC122956
VERSION     AC122956.4  GI:25137800
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 246991)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 246991)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 246991)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23666770.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXPC
                Center clone name: CH230-17E8
            -------------- Summary Statistics
```

```
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225722 bases at least Q40
                Consensus quality: 228759 bases at least Q30
                Consensus quality: 230490 bases at least Q20
                Estimated insert size: 231948; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   246991: contig of 246991 bp in length.
FEATURES             Location/Qualifiers
     source          1..246991
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-17E8"
     misc_feature    1..1165
                     /note="wgs_contig"
     misc_feature    complement(238413..239222)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH308764"
     misc_feature    244729..246991
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             19.3%;  Score 37;  DB 14;  Length 246991;
  Best Local Similarity   54.9%;  Pred. No. 1.5;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy         29 GACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAA 88
              || || |  | |  ||||| |    |||||    |   |   |    |||   |  | ||
Db       6776 GAAAAACCCATTTATGTCTAAGAGCTCTATGCAGATATCTAACTTCTAAAGGAACATAAA 6717

Qy         89 AACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
              ||  ||| | ||||  |  |||||  || |  ||| | ||||||  ||  | |   |  |
Db       6716 AATACTTTTCAAACATGTATTTTCTTAGTAGGATGAGGTAGTATACTTCTTTAACCCTAG 6657

Qy        149 GACTAAGCAAAAG 161
              ||||| ||| | |
Db       6656 GACTAGGCAGAGG 6644


RESULT 8
AL671518
LOCUS       AL671518               67395 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-315C6 on chromosome 6 Contains
            the 5' end of the WASF1 gene for WAS protein family member 1, the
            gene for pre-mRNA splicing factor 17 (PRP17, EH binding protein 3,
            EHB3, CDC40), a novel gene and a CpG island, complete sequence.
ACCESSION   AL671518
VERSION     AL671518.3  GI:18855253
KEYWORDS    HTG; CDC40; CpG island; EH binding protein; EHB3; KIAA0269;
            pre-mRNA splicing factor; PRP17; SCAR1; Verprolin-homologous
            protein; WAS protein family; WASF1; WAVE; WAVE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 67395)
  AUTHORS   Tracey,A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 21, 2002 this sequence version replaced gi:18491488.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-315C6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1..67395
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-315C6"
                     /clone_lib="RPCI-11.2"
     misc_feature    2000
                     /note="Clone_right_end: RP11-403I19"
     gene            join(complement(7668..7828),complement(6425..6569),
                     complement(4974..5013),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:<114860..114994))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-003"
     mRNA            join(complement(7668..7828),complement(6425..6569),
                     complement(4974..5013),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:<114860..114994))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-003"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: BG771905.1"
     gene            join(complement(7668..7828),complement(6425..6569),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:110062..110215),
                     complement(AL590009.10:108611..108728),
                     complement(AL590009.10:106941..107113),
                     complement(AL590009.10:104912..105091),
                     complement(AL590009.10:103122..103750),
                     complement(AL590009.10:101353..102213))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-002"
     mRNA            join(complement(7668..7828),complement(6425..6569),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
```

```
                     complement(AL590009.10:110062..110215),
                     complement(AL590009.10:108611..108728),
                     complement(AL590009.10:106941..107113),
                     complement(AL590009.10:104912..105091),
                     complement(AL590009.10:103122..103750),
                     complement(AL590009.10:101353..102213))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-002"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: AA515859.1 AA885055.1 AI750422.1
                     AL526626.1 AL702231.1 AU142666.1 AU159926.1 AW071606.1
                     AW673234.1 AW956469.1 AW963370.1 BG111973.1 BG176906.1
                     BG665254.1 BG806840.1 BM465114.1
                     match: cDNAs: AF134303.1 D87459.1"
     gene            join(complement(7668..7750),complement(6425..6569),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:<110062..110215))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-007"
     mRNA            join(complement(7668..7750),complement(6425..6569),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:<110062..110215))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-007"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: BE787088.1"
     gene            join(complement(7266..7476),complement(6425..6569),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:<110062..110215))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-001"
     gene            join(complement(7266..7476),complement(6425..6569),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:<114860..114994))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-006"
     mRNA            join(complement(7266..7476),complement(6425..6569),
                     complement(AL590549.9:11123..11220),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:<110062..110215))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-001"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: BI462105.1"
     mRNA            join(complement(7266..7476),complement(6425..6569),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:<114860..114994))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-006"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: BI461486.1"
     gene            join(complement(7266..7529),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:110062..110215),
                     complement(AL590009.10:108611..108728),
                     complement(AL590009.10:106941..107113),
                     complement(AL590009.10:104912..105091),
                     complement(AL590009.10:103122..103750),
                     complement(AL590009.10:101353..102213))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-004"
     mRNA            join(complement(7266..7529),
                     complement(AL590009.10:129003..129163),
                     complement(AL590009.10:114860..114994),
                     complement(AL590009.10:110062..110215),
                     complement(AL590009.10:108611..108728),
                     complement(AL590009.10:106941..107113),
```

```
                     complement(AL590009.10:104912. .105091),
                     complement(AL590009.10:103122. .103750),
                     complement(AL590009.10:101353. .102213))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-004"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: AA885055.1 AI154223.1 AI650710.1
                     AV664100.1 AW379048.1 AW673234.1 BF796356.1 BG111973.1
                     BG703130.1 BG778594.1 BG806840.1 BI544869.1 BI545690.1
                     BI545964.1 BI546470.1 BI550989.1 BI552818.1 BI989518.1
                     BM458302.1 H18215.1
                     match: cDNAs: AF134303.1 BC019019.1"
     gene            join(complement(7266. .7497),
                     complement(AL590549.9:11123. .11220),
                     complement(AL590009.10:129003. .129163),
                     complement(AL590009.10:114860. .114994),
                     complement(AL590009.10:110062. .110215),
                     complement(AL590009.10:<108611. .108728))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-005"
     mRNA            join(complement(7266. .7497),
                     complement(AL590549.9:11123. .11220),
                     complement(AL590009.10:129003. .129163),
                     complement(AL590009.10:114860. .114994),
                     complement(AL590009.10:110062. .110215),
                     complement(AL590009.10:<108611. .108728))
                     /gene="WASF1"
                     /locus_tag="RP11-181P4.1-005"
                     /product="WAS protein family, member 1"
                     /note="match: ESTs: BG719866.1 BI462401.1"
     gene            join(7968. .8007,8211. .8460,21009. .21095,29385. .29514,
                     35333. .35416,36911. .37050,38534. .38630,39960. .40099,
                     40913. .40987,43113. .43158,45529. .45630,47191. .47306,
                     47563. .47696,53994. .54070,56659. .56803,57781. .60044)
                     /gene="CDC40"
                     /locus_tag="RP1-71D21.3-005"
     mRNA            join(7968. .8007,8211. .8460,21009. .21095,29385. .29514,

  Query Match             19.0%;  Score 36.4;  DB 8;  Length 67395;
  Best Local Similarity   54.5%;  Pred. No. 2.4;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              || |||  | ||   || |  |||||| |||||| |       | |||| | | | |||
Db      15983 CAAAATAATTTCAATTTGGAATATTTTAAATACCCTTAATTGGAAAGATCTCATATCCCA 16042

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
              ||| | |    |||| | |||   | |  |  ||  ||| ||   ||  | ||| |
Db      16043 TATTATAAAGTGGTTCTAGCAACTATACACTATAAAATCCTTTGGGAAGATATGAATGTG 16102

Qy        155 GCAAAAGTGCTTGT 168
              || ||    | |||
Db      16103 GCTAAGCCTCATGT 16116


RESULT 9
AC154842
LOCUS       AC154842              317403 bp    DNA     linear   HTG 30-DEC-2004
DEFINITION  Mus musculus chromosome 17 clone RP23-55C11, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC154842 AC101665
VERSION     AC154842.1  GI:56900468
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 317403)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 317403)
```

```
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Dec 30, 2004 this sequence version replaced gi:45680671.

            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0055C11
            Drafting center: WIBR
            -----------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 312509 bases at least Q40
            Consensus quality: 313620 bases at least Q30
            Consensus quality: 314237 bases at least Q20
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1281: contig of 1281 bp in length
            *     1282     1381: gap of unknown length
            *     1382     2759: contig of 1378 bp in length
            *     2760     2859: gap of unknown length
            *     2860     6382: contig of 3523 bp in length
            *     6383     6482: gap of unknown length
            *     6483    13384: contig of 6902 bp in length
            *    13385    13484: gap of unknown length
            *    13485    19720: contig of 6236 bp in length
            *    19721    19820: gap of unknown length
            *    19821    26341: contig of 6521 bp in length
            *    26342    26441: gap of unknown length
            *    26442    34060: contig of 7619 bp in length
            *    34061    34160: gap of unknown length
            *    34161    53672: contig of 19512 bp in length
            *    53673    53772: gap of unknown length
            *    53773    65274: contig of 11502 bp in length
            *    65275    65374: gap of unknown length
            *    65375    76706: contig of 11332 bp in length
            *    76707    76806: gap of unknown length
            *    76807   102682: contig of 25876 bp in length
            *   102683   102782: gap of unknown length
            *   102783   137874: contig of 35092 bp in length
            *   137875   137974: gap of unknown length
            *   137975   173841: contig of 35867 bp in length
            *   173842   173941: gap of unknown length
            *   173942   317403: contig of 143462 bp in length.
FEATURES             Location/Qualifiers
     source          1. .317403
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /clone="RP23-55C11"
     misc_feature    1. .1281
                     /note="assembly_name:Contig41"
     gap             1282. .1381
                     /estimated_length=unknown
     misc_feature    1382. .2759
                     /note="assembly_name:Contig44"
```

```
     gap             2760. .2859
                     /estimated_length=unknown
     misc_feature    2860. .6382
                     /note="assembly_name:Contig47"
     gap             6383. .6482
                     /estimated_length=unknown
     misc_feature    6483. .13384
                     /note="assembly_name:Contig48"
     gap             13385. .13484
                     /estimated_length=unknown
     misc_feature    13485. .19720
                     /note="assembly_name:Contig49"
     gap             19721. .19820
                     /estimated_length=unknown
     misc_feature    19821. .26341
                     /note="assembly_name:Contig50"
     gap             26342. .26441
                     /estimated_length=unknown
     misc_feature    26442. .34060
                     /note="assembly_name:Contig51"
     gap             34061. .34160
                     /estimated_length=unknown
     misc_feature    34161. .53672
                     /note="assembly_name:Contig52"
     gap             53673. .53772
                     /estimated_length=unknown
     misc_feature    53773. .65274
                     /note="assembly_name:Contig53"
     gap             65275. .65374
                     /estimated_length=unknown
     misc_feature    65375. .76706
                     /note="assembly_name:Contig54"
     gap             76707. .76806
                     /estimated_length=unknown
     misc_feature    76807. .102682
                     /note="assembly_name:Contig55"
     gap             102683. .102782
                     /estimated_length=unknown
     misc_feature    102783. .137874
                     /note="assembly_name:Contig56"
     gap             137875. .137974
                     /estimated_length=unknown
     misc_feature    137975. .173841
                     /note="assembly_name:Contig57"
     gap             173842. .173941
                     /estimated_length=unknown
     misc_feature    173942. .317403
                     /note="assembly_name:Contig58"
ORIGIN

  Query Match             19.0%;  Score 36.4;  DB 14;  Length 317403;
  Best Local Similarity   54.5%;  Pred. No. 2.3;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         13 GGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCG 72
              |||  |   ||||||| ||| ||| |    || |||||     | ||  || || |||
Db      92985 GGCATTCTGGTGGCTTCACATGCAAATCACTGCCTGATGATAATGTTGCCAGTATCTTGC 93044

Qy         73 GGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTAT 132
                | | | | |||||| ||| | ||  |       |       ||| |   ||||  |||
Db      93045 ATGCATTGGGTGTGAATACCATAATGCAGTAGCCCTCCATTTTGAAGCCATCGCTGTTAT 93104

Qy        133 CATTGATGACAACA 146
              |  ||||  ||| |
Db      93105 CTCTGATTGCAAAA 93118


RESULT 10
AC140952
LOCUS       AC140952              170350 bp    DNA     linear   PRI 26-SEP-2003
DEFINITION  Pan troglodytes BAC clone RP43-22I12 from 7, complete sequence.
ACCESSION   AC140952
VERSION     AC140952.1  GI:28850254
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 170350)
  AUTHORS   Shah,N., Meyer,R. and Bielicki,L.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-22I12
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 170350)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Sequencing of Pan troglodytes
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 170350)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 170350)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT022I12
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .170350
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /map="7"
                     /clone="RP43-22I12"
                     /clone_lib="RPCI-43"
ORIGIN
```

```
  Query Match             18.6%;  Score 35.8;  DB 8;  Length 170350;
  Best Local Similarity   63.2%;  Pred. No. 3.7;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
               ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db      144944 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 145003

Qy          75 GAAATAGATGTGAAAACCCTTATAAAA 101
                 ||    | ||||||| |||| ||||
Db      145004 ATAAGCATTTTGAAAACACTTAAAAAA 145030


RESULT 11
AC092873
LOCUS       AC092873               171239 bp    DNA     linear   HTG 12-JUN-2002
DEFINITION  Pan troglodytes clone RP43-22I12, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC092873
VERSION     AC092873.3  GI:21392490
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 171239)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171239)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 171239)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 12, 2002 this sequence version replaced gi:15799603.
            ------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ------------- Project Information
                 Center project name: ani
                 Center clone name: 022I12

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 170538 bases at least Q40
                 Consensus quality: 170688 bases at least Q30
                 Consensus quality: 170863 bases at least Q20
                 Insert size: 152000; agarose-fp
                 Insert size: 168000; pulse-field-gel
                 Insert size: 170939; sum-of-contigs
                 Quality coverage: 30.12x in Q20 bases; agarose-fp
                 Quality coverage: 27.25x in Q20 bases; pulse-field-gel
                 Quality coverage: 26.78x in Q20 bases; sum-of-contigs
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1     82247: contig of 82247 bp in length
            *      82248     82347: gap of unknown length
            *      82348     83760: contig of 1413 bp in length
            *      83761     83860: gap of unknown length
            *      83861    168249: contig of 84389 bp in length
            *     168250    168349: gap of unknown length
            *     168350    171239: contig of 2890 bp in length.
FEATURES             Location/Qualifiers
     source          1..171239
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-22I12"
                     /clone_lib="RP43"
     misc_feature    1..82247
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1..62423
                     /note="clone overlaps with GenBank Accession Number
                     AC093139 clone RP43-22A1 (center project name anh)"
     misc_feature    61398..171239
                     /note="clone overlaps with GenBank Accession Number
                     AC098646 clone RP43-44G6 (center project name anj)"
     gap             82248..82347
                     /estimated_length=unknown
     misc_feature    82348..83760
                     /note="assembly_fragment"
     gap             83761..83860
                     /estimated_length=unknown
     misc_feature    83861..168249
                     /note="assembly_fragment"
     gap             168250..168349
                     /estimated_length=unknown
     misc_feature    168350..171239
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 14;  Length 171239;
  Best Local Similarity   63.2%;  Pred. No. 3.7;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
               ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db      145797 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 145856

Qy          75 GAAATAGATGTGAAAACCCTTATAAAA 101
                 ||    | ||||||| |||| ||||
Db      145857 ATAAGCATTTTGAAAACACTTAAAAAA 145883
```

```
RESULT 12
AC098646
LOCUS       AC098646               171391 bp    DNA     linear   HTG 13-JUN-2002
DEFINITION  Pan troglodytes clone RP43-44G6, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC098646
VERSION     AC098646.2  GI:21397305
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 171391)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171391)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 171391)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:16506408.
            ------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: anj
                 Center clone name: 044G06

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 170643 bases at least Q40
                 Consensus quality: 170930 bases at least Q30
                 Consensus quality: 171055 bases at least Q20
                 Insert size: 150000; agarose-fp
                 Insert size: 154000; pulse-field-gel
                 Insert size: 171091; sum-of-contigs
                 Quality coverage: 11.37x in Q20 bases; agarose-fp
                 Quality coverage: 11.07x in Q20 bases; pulse-field-gel
                 Quality coverage: 9.97x in Q20 bases; sum-of-contigs
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1     21570: contig of 21570 bp in length
            *      21571     21670: gap of unknown length
            *      21671     57729: contig of 36059 bp in length
            *      57730     57829: gap of unknown length
            *      57830    105820: contig of 47991 bp in length
            *     105821    105920: gap of unknown length
            *     105921    171391: contig of 65471 bp in length.
FEATURES             Location/Qualifiers
     source          1..171391
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-44G6"
                     /clone_lib="RP43"
     misc_feature    1..108807
                     /note="clone overlaps with GenBank Accession Number
                     AC092873 clone RP43-22I12 (center project name ani)"
     misc_feature    1..21570
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1..1027
                     /note="clone overlaps with GenBank Accession Number
                     AC093139 clone RP43-22A1 (center project name anh)"
     gap             21571..21670
                     /estimated_length=unknown
     misc_feature    21671..57729
                     /note="assembly_fragment"
     gap             57730..57829
                     /estimated_length=unknown
     misc_feature    57830..105820
                     /note="assembly_fragment"
     gap             105821..105920
                     /estimated_length=unknown
     misc_feature    105921..171391
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    138190..171391
                     /note="clone overlaps with GenBank Accession Number
                     AC092871 clone RP43-120E8 (center project name ank)"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 14;  Length 171391;
  Best Local Similarity   63.2%;  Pred. No. 3.7;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
               ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db       83381 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 83440

Qy          75 GAAATAGATGTGAAAACCCTTATAAAA 101
                 ||    | ||||||| |||| ||||
Db       83441 ATAAGCATTTTGAAAACACTTAAAAAA 83467


RESULT 13
AC151799/c
LOCUS       AC151799               183888 bp    DNA     linear   PRI 27-APR-2005
DEFINITION  Pan troglodytes BAC clone RP43-44D10 from chromosome 7, complete
            sequence.
ACCESSION   AC151799
VERSION     AC151799.2  GI:55168552
KEYWORDS    HTG.
```

```
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 183888)
  AUTHORS   Adams,S. and Bielicki,L.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-44D10
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 183888)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 183888)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2004) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 183888)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2005) Washington University School of Medicine,
            Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
            63108, USA
COMMENT     On Nov 2, 2004 this sequence version replaced gi:53386473.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT044D10


            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..183888
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /clone="RP43-44D10"
                     /clone_lib="RPCI-43"
     unsure          72798
                     /note="Unresolved bases"
ORIGIN

  Query Match             18.6%;  Score 35.8;  DB 8;  Length 183888;
```

```
  Best Local Similarity   63.2%;  Pred. No. 3.7;
  Matches   55;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         15 CCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGG 74
              ||| | | ||| |  | ||| || || || | | |||  ||||| || || || ||
Db     100188 CCAGGCACTGGGTACATAAGTATCATATTTTTTAATTTCTCTATGTTTAAAACATTTTAT 100129

Qy         75 GAAATAGATGTGAAAACCCTTATAAAA 101
                ||    | ||||||| |||| ||||
Db     100128 ATAAGCATTTTGAAAACACTTAAAAAA 100102


RESULT 14
AC149460/c
LOCUS       AC149460              181622 bp    DNA     linear   HTG 08-JUL-2004
DEFINITION  Papio anubis clone RP41-441E1, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC149460
VERSION     AC149460.2  GI:50057884
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 181622)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
            Coleman,H., Daki,N., Engle,J., Guan,X., Gupta,J., Haghighi,P.,
            Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C.,
            Karlins,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q.,
            Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,E.H.,
            Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C.,
            Park,M., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N.,
            Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D.,
            Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181622)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 181622)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 8, 2004 this sequence version replaced gi:48057351.
            -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: hbl
                Center clone name: 441E01

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
```

```
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180653 bases at least Q40
            Consensus quality: 180918 bases at least Q30
            Consensus quality: 181167 bases at least Q20
            Insert size: 180000; agarose-fp
            Insert size: 181322; sum-of-contigs
            Quality coverage: 10.99x in Q20 bases; agarose-fp
            Quality coverage: 10.91x in Q20 bases; sum-of-contigs
            --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1     2622: contig of 2622 bp in length
          *     2623     2722: gap of unknown length
          *     2723   132243: contig of 129521 bp in length
          *   132244   132343: gap of unknown length
          *   132344   162447: contig of 30104 bp in length
          *   162448   162547: gap of unknown length
          *   162548   181622: contig of 19075 bp in length.
FEATURES             Location/Qualifiers
     source          1..181622
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-441E1"
                     /clone_lib="RP41"
                     /note="BAC resource: http://bacpac.chori.org/"
     misc_feature    1..2622
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             2623..2722
                     /estimated_length=unknown
     misc_feature    2723..132243
                     /note="assembly_fragment"
     gap             132244..132343
                     /estimated_length=unknown
     misc_feature    132344..162447
                     /note="assembly_fragment"
     misc_feature    132495..181622
                     /note="clone overlaps with GenBank Accession Number
                     AC150199 clone RP41-189I7 (center project name hbk)"
     gap             162448..162547
                     /estimated_length=unknown
     misc_feature    162548..181622
                     /note="assembly_fragment
                     missing T7 clone end on 3' end of insert"
ORIGIN

  Query Match             18.2%;  Score 35;  DB 14;  Length 181622;
  Best Local Similarity   53.2%;  Pred. No. 6.6;
  Matches   74;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAA 101
               || |||  || | ||  | | |  | || |  | ||||  || ||   | |   |||||
Db       73060 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTTAGCAGAATAAAG 73001

Qy         102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |     |||  ||| |||  ||    ||  || ||| |     | ||   ||||||   |
Db       73000 CCTATTTTTCAGCATAAATATTGCTTACCATGATTAAAATAGAAATCAGCTAAGCTTGAT 72941

Qy         162 TGCTTGTCCCCTGACCCAA 180
               | |||  |   ||  ||||
Db       72940 TTCTTTACTGGTGGACCAA 72922
```

```
RESULT 15
AC116068
LOCUS       AC116068              231844 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-66D24, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC116068
VERSION     AC116068.5  GI:30579201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 231844)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231844)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231844)
  AUTHORS   Rat Genome Sequencing Consortium.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24818872.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMWC
                Center clone name: CH230-66D24
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 226512 bases at least Q40
                Consensus quality: 227588 bases at least Q30
                Consensus quality: 228638 bases at least Q20
                Estimated insert size: 242654; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   229543: contig of 229543 bp in length
            *   229544   229643: gap of unknown length
            *   229644   230670: contig of 1027 bp in length
            *   230671   230770: gap of unknown length
            *   230771   231844: contig of 1074 bp in length.
FEATURES             Location/Qualifiers
     source          1..231844
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-66D24"
     misc_feature    1..2343
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(2941..3750)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH264906"
     misc_feature    182235..225520
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH264909"
     misc_feature    227512..229543
                     /note="wgs_end_extension
                     clone_end:Sp6"
     gap             229544..229643
                     /estimated_length=unknown
     gap             230671..230770
                     /estimated_length=unknown
ORIGIN

  Query Match             18.2%;  Score 35;  DB 14;  Length 231844;
  Best Local Similarity   55.3%;  Pred. No. 6.5;
  Matches   68;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy          14 GCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGG 73
               ||| | |||||  |    ||||| |      | |  |   | | ||| ||| | ||||
Db        9334 GCCTTTAAGTGTTTAATAAAGCAAACCACACTGTCTTATTTGTCTTTACAACAGCTTCAT 9393

Qy          74 GGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATC 133
                |||||  | || | |  || || ||  || |||||||  || |   |||   | |
Db        9394 TGAAATGTAAGTCACATGCCATAAAATTCGGGGGTTTTTCCACACTTATGAATTCGGTGG 9453

Qy         134 ATT 136
               |||
Db        9454 ATT 9456

Search completed: March  7, 2006, 12:15:05
Job time : 1788.23 secs
```

```
                    GenCore version 5.1.7
           Copyright (c) 1993 - 2006  Biocceleration Ltd.


OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 10:55:04 ; Search time 453.383 Seconds
                                           (without alignments)
                                           2822.383 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters:      9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_21:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002as:*
                7:  geneseqn2002bs:*
                8:  geneseqn2003as:*
                9:  geneseqn2003bs:*
                10:  geneseqn2003cs:*
                11:  geneseqn2003ds:*
                12:  geneseqn2004as:*
                13:  geneseqn2004bs:*
                14:  geneseqn2005s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result              Query
   No.    Score    Match Length DB  ID                         Description
-------------------------------------------------------------------------------
     1      192    100.0    600  6  ABK52324                   Abk52324 DNA encod
c    2    175.2     91.2 156638  6  ABQ81850                   Abq81850 Bifidobac
c    3     33.4     17.4  50000  3  AAA64140                   Aaa64140 Nucleotid
     4     32.6     17.0    278  3  AAC27526                   Aac27526 Human sec
c    5     32.4     16.9   2717  4  ABL21958                   Abl21958 Drosophil
     6       32     16.7 110000  6  ABQ69245_15                Continuation (16 o
c    7       32     16.7 110000  6  ABQ67195_2                 Continuation (3 of
c    8     31.2     16.2  18272  4  AAK69447                   Aak69447 Human imm
c    9     31.2     16.2  18272  4  AAL06181                   Aal06181 Human rep
c   10     31.2     16.2  18272  4  ABL98746                   Abl98746 Human tes
c   11     31.2     16.2  26167 14   ADX83784                   Adx83784 Human aci
    12     30.6     15.9    466 13   ACN50973                   Acn50973 Cotton an
c   13     30.6     15.9    549 12   ADJ44041                   Adj44041 Plant cDN
    14     30.4     15.8    583  5  ABV51855                   Abv51855 Human pro
    15     30.4     15.8   1569  4  AAH52605                   Aah52605 S. epider
    16     30.4     15.8   3204  4  AAH54121                   Aah54121 S. epider
    17     30.2     15.7    515  4  AAK37564                   Aak37564 Human bon
    18     30.2     15.7   2000 11   ACL38062                   Acl38062 Rice stre
    19     30.2     15.7 110000 11   ACN44014_2                 Continuation (3 of
```

```
c   20       30     15.6   1332 13   ADS47100                   Ads47100 Bacterial
    21       30     15.6   7017  4  AAS32482                   Aas32482 Human gen
    22       30     15.6  25519 11   ACN44652                   Acn44652 Mouse gen
c   23     29.8     15.5  10303 13   ADI67110                   Adi67110 Novel Lac
    24     29.8     15.5  10304 13   ADI67125                   Adi67125 Novel Lac
c   25     29.8     15.5 115829 13   ABD33448                   Abd33448 Human can
c   26     29.6     15.4   1824 14   AEB26996                   Aeb26996 Pinus rad
c   27     29.6     15.4   1824 14   AEB26999                   Aeb26999 Pinus rad
    28     29.6     15.4   2000  8  ADA71889                   Ada71889 Rice gene
    29     29.6     15.4   2000 11   ACL35459                   Acl35459 Rice stre
    30     29.6     15.4   2510 13   ADX49899                   Adx49899 Plant ful
    31     29.6     15.4 110000 14   AEB42401_28                Continuation (29 o
    32     29.6     15.4 110000 14   AEB42737_14                Continuation (15 o
c   33     29.6     15.4 200622 14   AEB39167                   Aeb39167 L. pneumo
    34     29.6     15.4 260027 11   ACN44046                   Acn44046 Human gen
    35     29.4     15.3    487  9  ACH25462                   Ach25462 Human adu
    36     29.4     15.3    768 10   ADH85446                   Adh85446 Enterococ
c   37     29.4     15.3    867 13   ADS57789                   Ads57789 Bacterial
c   38     29.4     15.3   2000 10   ACC60804                   Acc60804 Gene sequ
c   39     29.4     15.3   2000 10   ADK62093                   Adk62093 Disease t
c   40     29.4     15.3   2414  4  ABL22762                   Abl22762 Drosophil
c   41     29.4     15.3   2537  4  ABL11638                   Abl11638 Drosophil
c   42     29.4     15.3   2719  4  ABL06226                   Abl06226 Drosophil
c   43     29.4     15.3   3612 13   ADT47758                   Adt47758 Bacterial
    44     29.4     15.3   4088  4  ABL03916                   Abl03916 Drosophil
c   45     29.2     15.2    836  4  AAK83773                   Aak83773 Human imm


                                  ALIGNMENTS


RESULT 1
ABK52324
ID   ABK52324 standard; DNA; 600 BP.
XX
AC   ABK52324;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding cancer gene therapy associated protein.
XX
KW   Solid cancer; anaerobic bacteria; antitumour active protein; cancer;
KW   gene therapy; gene; ds.
XX
OS   Bifidobacterium longum.
XX
FH   Key             Location/Qualifiers
FT   CDS             193. .474
FT                   /*tag=  a
FT                   /product= "Cancer gene therapy protein"
XX
PN   JP2002097144-A.
XX
PD   02-APR-2002.
XX
PF   21-SEP-2001; 2001JP-00290187.
XX
PR   21-SEP-2000; 2000JP-00287688.
XX
PA   (AMAN/) AMANO A.
PA   (FUJI/) FUJIMORI M.
XX
DR   WPI; 2002-448201/48.
DR   P-PSDB; AAU96807.
XX
PT   Solid cancer therapy with anaerobic bacteria of Bifidobacterium sp. by
PT   tumor tissue specific delivery of a DNA encoding for an antitumor active
PT   protein or its precursor.
XX
PS   Claim 10; Page 16; 21pp; Japanese.
XX
CC   The invention describes a method of treating a solid cancer with
CC   anaerobic bacteria by site specific delivery of DNA encoding an
```

```
CC   antitumour active protein or its precursor. This sequence encodes a
CC   cancer gene therapy associated protein
XX
SQ   Sequence 600 BP; 132 A; 165 C; 161 G; 142 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 192;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-57;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
Db        181 GAAGGATGCTTT 192


RESULT 2
ABQ81850/c
ID   ABQ81850 standard; DNA; 156638 BP.
XX
AC   ABQ81850;
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1106.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
OS   Synthetic.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-00102050.
XX
PR   30-JAN-2001; 2001EP-00102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT   a probe or primer for detecting and/or identifying Bifidobacterium longum
PT   in a biological sample.
XX
PS   Disclosure; SEQ ID NO 1106; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC   fusion protein, comprising a sequence selected from 1097 sequences given
CC   in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC   heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC   activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC   a probe) is useful for the detection and/or identification of
CC   Bifidobacterium longum in a biological sample. A carrier containing the
CC   lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
CC   used for preventing and/or treating diarrhoea brought about by pathogenic
CC   bacteria and/or rotavirus. The carrier is a food composition selected
CC   from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC   products, ice-creams, fermented cereal based products, milk based
CC   powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification. N.B. The sequence data for this patent is not represented
CC   in the printed specification but is based on sequence information
CC   supplied by the European Patent Office
XX
SQ   Sequence 156638 BP; 32098 A; 46491 C; 46415 G; 31634 T; 0 U; 0 Other;

  Query Match             91.2%;  Score 175.2;  DB 6;  Length 156638;
  Best Local Similarity   97.4%;  Pred. No. 3.5e-50;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142164 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 142105

Qy         61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
              ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142104 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 142045

Qy        119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142044 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 141985

Qy        179 AAGAAGGATGCTTT 192
              ||||||||||||||
Db     141984 AAGAAGGATGCTTT 141971


RESULT 3
AAA64140/c
ID   AAA64140 standard; DNA; 50000 BP.
XX
AC   AAA64140;
XX
DT   20-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of a beta-tubulin antigen.
XX
KW   Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;
KW   chronic ear disease; autoimmune disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200050593-A1.
XX
PD   31-AUG-2000.
XX
PF   25-FEB-2000; 2000WO-US004795.
XX
PR   25-FEB-1999;   99US-0121549P.
XX
PA   (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI   Yoo TJ;
XX
DR   WPI; 2000-558400/51.
XX
PT   New beta-tubulin antigen in the membranous structure of the inner ear,
PT   reactive with antibodies of patients with Meniere's disease, for
PT   diagnosing Meniere's disease and distinguishing this disease from other
PT   autoimmune ear diseases.
XX
```

```
PS   Claim 3; Page 74-97; 115pp; English.
XX
CC   The present sequence encodes a beta-tubulin antigen. The protein is an
CC   antigen of the membranous structure of the inner ear protein, and is
CC   reactive with antibodies from patients having Meniere's disease.
CC   Meniere's disease is a chronic ear disease with unknown etiology. Serum
CC   from patients suffering from this disease contain autoantibodies against
CC   a 30 kDa cochlear protein antigen. The disease is believed to be an
CC   autoimmune disease. The beta-tubulin antigen is useful as a target
CC   substance in diagnosing or detecting Meniere's disease and in
CC   distinguishing this disease from other autoimmune ear diseases
XX
SQ   Sequence 50000 BP; 17281 A; 9480 C; 8791 G; 14448 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 33.4;  DB 3;  Length 50000;
  Best Local Similarity   52.5%;  Pred. No. 3.2;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGAAATAGATGTGAAAACCCTTATAAAA 101
               || |||  || | ||  | | |  | || |  | ||||  || ||   | |   || ||
Db       26376 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTCAGCAGAATGAAG 26317

Qy         102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |     |||  ||| ||| ||    |   || ||| |     | ||   |||||||  |
Db       26316 CCTATTTTTCAGCATAAATATTGCTTGCCATGATTAAAATAGAAATCAGCTAAGCATGAT 26257

Qy         162 TGCTTGTCCCCTGACCCAA 180
               | |||  |   ||   ||||
Db       26256 TTCTTTACTGGTGGACCAA 26238


RESULT 4
AAC27526
ID   AAC27526 standard; cDNA; 278 BP.
XX
AC   AAC27526;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 31601.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 31601; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 278 BP; 82 A; 44 C; 43 G; 108 T; 0 U; 1 Other;

  Query Match             17.0%;  Score 32.6;  DB 3;  Length 278;
  Best Local Similarity   51.0%;  Pred. No. 0.63;
  Matches   77;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          18 TGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAA 77
               | | |||| |||       |  |   || | |     || | |||||||  || |||||
Db          99 TCATGTGGATTGTTTTCTGTCTTTAGGTATTAAAATGCTGTATTCAATAATTTKGGGGAC 158

Qy          78 ATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTG 137
               ||| |||| |   | | |||| |  |  |  | ||   |  |||| |   |    || ||
Db         159 ATATATGTCATTTCACATATATACAGATGTGTATCAGGGTGACATTCCAAAAGTGCACTG 218

Qy         138 ATGACAACATGGACTAAGCAAAAGTGCTTGT 168
                |||      ||  || | |   ||  || |
Db         219 CTGAGTTAGAGGTATATGTATTTGTAATTTT 249


RESULT 5
ABL21958/c
ID   ABL21958 standard; DNA; 2717 BP.
XX
AC   ABL21958;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 17347.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 17347; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
```

```
XX
SQ   Sequence 2717 BP; 835 A; 555 C; 503 G; 824 T; 0 U; 0 Other;

  Query Match             16.9%;  Score 32.4;  DB 4;  Length 2717;
  Best Local Similarity   56.6%;  Pred. No. 2;
  Matches   60;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          39 ATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATA 98
               ||  |  | ||| | |  |  | |||||| |      |  || || |  | |  |||| |
Db         369 ATAGTTACAGATACATAAAAATACAATACATAGCAAAAGGTACATATATAGAGACTTAAA 310

Qy          99 AAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
               | ||||  |||||| | |||  || |  ||  ||  ||   |||||
Db         309 ATACGCTTGTTTTCACTGAATTATTCAATACGATTTTTAGAGACAA 264


RESULT 6
ABQ69245_15
Continuation (16 of 31) of ABQ69245 from base 1500001 (Listeria innocua DNA sequence #68
WP Sequence split into 31 fragments  LOCUS ABQ69245 Accession Abq69245
WP     Fragment Name      Begin      End
WP     ABQ69245_00            1      110000
WP     ABQ69245_01       100001      210000
WP     ABQ69245_02       200001      310000
WP     ABQ69245_03       300001      410000
WP     ABQ69245_04       400001      510000
WP     ABQ69245_05       500001      610000
WP     ABQ69245_06       600001      710000
WP     ABQ69245_07       700001      810000
WP     ABQ69245_08       800001      910000
WP     ABQ69245_09       900001     1010000
WP     ABQ69245_10      1000001     1110000
WP     ABQ69245_11      1100001     1210000
WP     ABQ69245_12      1200001     1310000
WP     ABQ69245_13      1300001     1410000
WP     ABQ69245_14      1400001     1510000
WP     ABQ69245_15      1500001     1610000
WP     ABQ69245_16      1600001     1710000
WP     ABQ69245_17      1700001     1810000
WP     ABQ69245_18      1800001     1910000
WP     ABQ69245_19      1900001     2010000
WP     ABQ69245_20      2000001     2110000
WP     ABQ69245_21      2100001     2210000
WP     ABQ69245_22      2200001     2310000
WP     ABQ69245_23      2300001     2410000
WP     ABQ69245_24      2400001     2510000
WP     ABQ69245_25      2500001     2610000
WP     ABQ69245_26      2600001     2710000
WP     ABQ69245_27      2700001     2810000
WP     ABQ69245_28      2800001     2910000
WP     ABQ69245_29      2900001     3010000
WP     ABQ69245_30      3000001     3011208

  Query Match             16.7%;  Score 32;  DB 6;  Length 110000;
  Best Local Similarity   50.7%;  Pred. No. 14;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
               ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db       29778 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 29837

Qy          95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
                || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db       29838 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 29897

Qy         155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
               ||    ||    |   ||  | | ||  | ||
Db       29898 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 29929


RESULT 7
ABQ67195_2/c
Continuation (3 of 5) of ABQ67195 from base 200001 (Listeria innocua contig DNA sequence
WP Sequence split into 5 fragments  LOCUS ABQ67195 Accession Abq67195
WP     Fragment Name      Begin     End
WP     ABQ67195_0             1     110000
WP     ABQ67195_1        100001     210000
WP     ABQ67195_2        200001     310000
WP     ABQ67195_3        300001     410000
WP     ABQ67195_4        400001     495269

  Query Match             16.7%;  Score 32;  DB 6;  Length 110000;
  Best Local Similarity   50.7%;  Pred. No. 14;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy          35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
               ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db       29177 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 29118

Qy          95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
                || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db       29117 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 29058

Qy         155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
               ||    ||    |   ||  | | ||  | ||
Db       29057 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 29026


RESULT 8
AAK69447/c
ID   AAK69447 standard; DNA; 18272 BP.
XX
AC   AAK69447;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24259.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
```

```
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
```

```
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 24259; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 12;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
```

```
              || | | | |||  ||||||  |  || |   |  ||| |  ||||| |  || |  ||
Db      17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 9
AAL06181/c
ID   AAL06181 standard; DNA; 18272 BP.
XX
AC   AAL06181;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8869.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
```

```
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 8869; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 12;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 10
ABL98746/c
ID   ABL98746 standard; DNA; 18272 BP.
XX
AC   ABL98746;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 3398.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
```

```
PF   17-JAN-2001; 2001WO-US001329.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
```

```
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483232/52.
XX
PT   Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT   for preventing, diagnosing and/or treating testicular cancer.
XX
PS   Disclosure; SEQ ID NO 3398; 766pp; English.
XX
CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention
XX
SQ   Sequence 18272 BP; 3601 A; 5274 C; 5400 G; 3997 T; 0 U; 0 Other;

  Query Match             16.2%;  Score 31.2;  DB 4;  Length 18272;
  Best Local Similarity   60.7%;  Pred. No. 12;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      17088 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17029

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17028 AACACCCATGAGAACAGGGACTCA 17005


RESULT 11
ADX83784/c
ID   ADX83784 standard; DNA; 26167 BP.
XX
AC   ADX83784;
XX
DT   05-MAY-2005  (first entry)
XX
DE   Human acid alpha glucosidase gene.
XX
KW   Recombinant DNA; transgenic animal; milk; alpha glucosidase; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            777. .1184
FT                   /*tag=  a
FT                   /number= 1
FT   intron          1185. .3848
FT                   /*tag=  b
FT                   /number= 1
FT   misc_feature    3281. .3541
FT                   /*tag=  c
FT                   /note= "Alu repeat"
FT   exon            3849. .4426
FT                   /*tag=  d
FT                   /number= 2
FT   misc_signal     3881. .3883
FT                   /*tag=  e
FT                   /note= "Translational start codon"
FT   intron          4427. .5042
FT                   /*tag=  f
FT                   /number= 2
FT   exon            5043. .5188
FT                   /*tag=  g
FT                   /number= 3
FT   intron          5189. .6851
FT                   /*tag=  h
FT                   /number= 3
FT   misc_feature    5940. .6217
FT                   /*tag=  i
FT                   /note= "Alu repeat"
FT   exon            6851. .7016
FT                   /*tag=  j
FT                   /number= 4
FT   intron          7017. .7100
```

```
FT                   /*tag=  k
FT                   /number= 4
FT   exon            7101. .7197
FT                   /*tag=  l
FT                   /number= 5
FT   intron          7198. .7590
FT                   /*tag=  m
FT                   /number= 5
FT   exon            7591. .7710
FT                   /*tag=  n
FT                   /number= 6
FT   intron          7711. .7789
FT                   /*tag=  o
FT                   /number= 6
FT   exon            7790. .7908
FT                   /*tag=  p
FT                   /number= 7
FT   intron          7909. .7997
FT                   /*tag=  q
FT                   /number= 7
FT   exon            7998. .8129
FT                   /*tag=  r
FT                   /number= 8
FT   intron          8130. .9238
FT                   /*tag=  s
FT                   /number= 8
FT   exon            9239. .9349
FT                   /*tag=  t
FT                   /number= 9
FT   intron          9350. .10020
FT                   /*tag=  u
FT                   /number= 9
FT   exon            10021. .10134
FT                   /*tag=  v
FT                   /number= 10
FT   intron          10135. .10234
FT                   /*tag=  w
FT                   /number= 10
FT   exon            10235. .10319
FT                   /*tag=  x
FT                   /number= 11
FT   intron          10320. .11275
FT                   /*tag=  y
FT                   /number= 11
FT   exon            11276. .11393
FT                   /*tag=  z
FT                   /number= 12
FT   intron          11394. .11870
FT                   /*tag=  aa
FT                   /number= 12
FT   exon            11871. .12004
FT                   /*tag=  ab
FT                   /number= 13
FT   intron          12005. .12168
FT                   /*tag=  ac
FT                   /number= 13
FT   exon            12169. .12320
FT                   /*tag=  ad
FT                   /number= 14
FT   intron          12321. .12510
FT                   /*tag=  ae
FT                   /number= 14
FT   exon            12511. .12659
FT                   /*tag=  af
FT                   /number= 15
FT   intron          12660. .16257
FT                   /*tag=  ag
FT                   /number= 15
FT   misc_feature    13030. .13308
FT                   /*tag=  ah
FT                   /note= "Alu repeat"
FT   exon            16258. .16399
FT                   /*tag=  ai
```

```
FT                   /number= 16
FT   intron          16400. .16907
FT                   /*tag=  aj
FT                   /number= 16
FT   exon            16908. .17057
FT                   /*tag=  ak
FT                   /number= 17
FT   intron          17058. .17504
FT                   /*tag=  al
FT                   /number= 17
FT   exon            17505. .17669
FT                   /*tag=  am
FT                   /number= 18
FT   intron          17670. .17964
FT                   /*tag=  an
FT                   /number= 18
FT   exon            17965. .18117
FT                   /*tag=  ao
FT                   /number= 19
FT   intron          18118. .18584
FT                   /*tag=  ap
FT                   /number= 19
FT                   /cons_splice= (5'site:NO, 3'site:YES)
FT   exon            18585. .19194
FT                   /*tag=  aq
FT                   /number= 20
FT   misc_signal     18642. .18644
FT                   /*tag=  ar
FT                   /note= "Translational stop codon"
FT   polyA_signal    19172. .19177
FT                   /*tag=  as
FT   misc_feature    19510. .19540
FT                   /*tag=  at
FT                   /note= "CT-rich region"
FT   misc_feature    20260. .20384
FT                   /*tag=  au
FT                   /note= "Half Alu repeat"
FT   misc_feature    20385. .20675
FT                   /*tag=  av
FT                   /note= "Alu repeat"
FT   misc_feature    20685. .20834
FT                   /*tag=  aw
FT                   /note= "Half Alu repeat"
FT   misc_feature    20845. .21132
FT                   /*tag=  ax
FT                   /note= "Alu repeat"
FT   misc_feature    22000. .22320
FT                   /*tag=  ay
FT                   /note= "Alu repeat"
FT   misc_feature    22675. .22964
FT                   /*tag=  az
FT                   /note= "Alu repeat"
FT   misc_feature    24649. .24923
FT                   /*tag=  ba
FT                   /note= "Alu repeat"
XX
PN   US6858425-B1.
XX
PD   22-FEB-2005.
XX
PF   03-DEC-1999;   99US-00454466.
XX
PR   04-DEC-1998;   98US-0110850P.
PR   02-MAR-1999;   99US-0122550P.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Heus JJ;
XX
DR   WPI; 2005-170882/18.
XX
PT   New recombinant polynucleotide comprises an alpha S1-casein promoter or
PT   an alpha glucosidase promoter, useful for designing transgenes for
```

```
PT   expression of human acid alpha glucosidase in the milk of transgenic
PT   animals.
XX
PS   Disclosure; SEQ ID NO 1; 49pp; English.
XX
CC   The invention relates to a recombinant polynucleotide comprising an alpha
CC   S1-casein promoter or an alpha glucosidase promoter operably linked to a
CC   heterologous coding sequence. The recombinant polynucleotide is useful
CC   for designing transgenes for expression of human acid alpha glucosidase
CC   in the milk of transgenic animals. It is also useful for the design of

  Query Match             16.2%;  Score 31.2;  DB 14;  Length 26167;
  Best Local Similarity   60.7%;  Pred. No. 14;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  || 
Db      17958 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17899

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17898 AACACCCATGAGAACAGGGACTCA 17875


RESULT 12
ACN50973
ID   ACN50973 standard; cDNA; 466 BP.
XX
AC   ACN50973;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton androecium tissue EST Clone ID: LIB3828-002-Q1-N6-F6, SEQ:5754.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
KW   variety Nucotton33B; library LIB3828; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 5754; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
```

```
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety Nucotton33B androecium tissue cDNA library (LIB3828). The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 466 BP; 138 A; 105 C; 81 G; 142 T; 0 U; 0 Other;

  Query Match             15.9%;  Score 30.6;  DB 13;  Length 466;
  Best Local Similarity   58.1%;  Pred. No. 4;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||   |   |  | || ||   | |||| | |   | | |  |  || ||||||
Db         27 ATATATATCATATGTCTTTCATTAACTTCTGGTTCTGGACTACATAATCCTTATTATCAT 86

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
              |||| | | |  |||  || ||| ||  |||||
Db         87 TGATCATACCCGGGAGGAACCAAGAGCCCTTGT 119


RESULT 13
ADJ44041/c
ID   ADJ44041 standard; cDNA; 549 BP.
XX
AC   ADJ44041;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #5041.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
```

```
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P,  Moughamer T,  Briggs SP,  Cooper B,  Glazebrook J;
PI   Goff SA,  Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Example 13; SEQ ID NO 5041; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 549 BP; 136 A; 143 C; 88 G; 177 T; 0 U; 5 Other;

  Query Match             15.9%;  Score 30.6;  DB 12;  Length 549;
  Best Local Similarity   51.9%;  Pred. No. 4.3;
  Matches   69;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         48 GATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGG 107
              || || | |   ||||| ||    |  |||| || | |  ||   |||    ||   | |
Db        235 GAATCATTTGAGTTCAAGACTGCTGAAGAAAAAGTTATCGAAGGTCTTGACCAAGCAGTG 176

Qy        108 TTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTG 167
               |  ||  |||  | |      |  | ||| ||||||      || |  |    | ||||
Db        175 ATGACGATGAATAAGGGAGAGATTGCTTTGGTGACAATTCCACCTGAATATGCCTTCTTG 116

Qy        168 TCCCCTGACCCAA 180
              ||| ||||  |||
Db        115 TCCACTGAATCAA 103


RESULT 14
ABV51855
ID   ABV51855 standard; cDNA; 583 BP.
XX
AC   ABV51855;
XX
DT   17-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 51846.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
```

```
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 10067-10068; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 583 BP; 210 A; 79 C; 104 G; 188 T; 0 U; 2 Other;

  Query Match             15.8%;  Score 30.4;  DB 5;  Length 583;
  Best Local Similarity   52.9%;  Pred. No. 5.2;
  Matches   64;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              | || ||||   | |||     ||||||   || |   |||| |  |||    |  | ||
Db         18 AGTCTTCTAGNATTAATTAACGCGGGGATTGAGTTCATAAAAGCTATATTTCTCATGTGT 77

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
              ||  |||||||  | |   | ||||||   | | |    ||   ||   |||    || |
Db         78 TTGGGCAGAAATTTTCAAGAATATCATATTTCAGATTTAGGGAAAACTCAAAAGTTTTAT 137

Qy        169 C 169
              |
Db        138 C 138


RESULT 15
AAH52605
ID   AAH52605 standard; DNA; 1569 BP.
XX
AC   AAH52605;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis open reading frame nucleotide sequence SEQ ID NO:603.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW   endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
```

```
XX
PF   09-NOV-2000; 2000WO-US030782.
XX
PR   09-NOV-1999;   99US-0164258P.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
DR   P-PSDB; AAG81755.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis.
XX
PS   Claim 8; Page 195; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC   and (II) can have antibacterial activity and therefore can be used in
CC   vaccination. The nucleic acids (I) may be used to produce the S.
CC   epidermidis polypeptides (II) via the production of vectors containing
CC   them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC   represent oligonucleotide sequences and primers which are used in the
CC   exemplification of the present invention. N.B. The present invention
CC   specifically claims all the polynucleotide sequences given in the
CC   sequence listing of the present specification, however the sequence
CC   listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC   in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC   for SEQ ID NO:4455 to 4464
XX
SQ   Sequence 1569 BP; 616 A; 186 C; 274 G; 493 T; 0 U; 0 Other;

  Query Match             15.8%;  Score 30.4;  DB 4;  Length 1569;
  Best Local Similarity   57.3%;  Pred. No. 8.1;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db         87 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 146

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        147 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 182

Search completed: March  7, 2006, 12:31:05
Job time : 457.383 secs
```

```
                    GenCore version 5.1.7
           Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 11:55:06 ; Search time 2362.02 Seconds
                                           (without alignments)
                                           3803.154 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters:      82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_est3:*
                4:  gb_htc:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_est7:*
                9:  gb_gss1:*
                10:  gb_gss2:*
                11:  gb_gss3:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result              Query
  No.   Score   Match Length DB  ID                    Description
-------------------------------------------------------------------------
      1   40.8   21.2    673  9  CE144436              CE144436 tigr-gss-
c     2   36.8   19.2    826 10   DU019646              DU019646 276061 To
c     3   36.4   19.0    749  9  BH039236              BH039236 RPCI-24-2
      4     35   18.2   1275  8  DN671770              DN671770 CFW64-F08
      5   34.6   18.0    852  2  BG542199              BG542199 602571637
      6   34.4   17.9    590  9  BH378221              BH378221 AG-ND-119
c     7   34.2   17.8    600  9  AZ964976              AZ964976 2M0234D24
      8   34.2   17.8    619  3  BP519798              BP519798 BP519798
      9   34.2   17.8    651  8  DN244174              DN244174 ACAE-aaa2
     10   34.2   17.8    878  7  CR281388              CR281388 CR281388
c    11   33.8   17.6    945 11   CNS079HE              AL435288 T3 end of
c    12   33.6   17.5    845 11   CNS0757U              AL429760 clone BA0
c    13   33.4   17.4    463 10   CW393812              CW393812 fsbb001f0
     14   33.4   17.4    494  6  CA394392              CA394392 cs51a07.y
     15   33.4   17.4    728  9  BH885133              BH885133 hw61c11.b
     16   33.2   17.3    449  7  CN673288              CN673288 A0932B01-
     17   33.2   17.3    467  6  CF166270              CF166270 B0768E01-
     18   33.2   17.3    493  7  CK391130              CK391130 K0827F01-
     19   33.2   17.3    501  6  CD552026              CD552026 B0337C08-
     20   33.2   17.3    501  6  CF165236              CF165236 B0753H06-
     21   33.2   17.3    503  6  CF168303              CF168303 B0797D04-
     22   33.2   17.3    538  6  CF162296              CF162296 B0709D06-
```

```
     23   33.2   17.3    551  7  CK391037              CK391037 K0826D11-
     24   33.2   17.3    552  7  CK390292              CK390292 K0817C11-
     25   33.2   17.3    552  7  CN690065              CN690065 E0288A10-
     26   33.2   17.3    555  6  CA885628              CA885628 B0120F01-
     27   33.2   17.3    555  6  CD553021              CD553021 B0353C01-
     28   33.2   17.3    557  6  CD553973              CD553973 B0369F09-
     29   33.2   17.3    572  7  CN682591              CN682591 E0166H03-
     30   33.2   17.3    579  6  CD540845              CD540845 B0226C02-
     31   33.2   17.3    589  7  CN676248              CN676248 A0977D08-
     32   33.2   17.3    596  7  CN681180              CN681180 E0147C03-
c    33   33.2   17.3    610  9  CC007613              CC007613 PUEDB53TD
c    34   33.2   17.3    611  9  AQ017410              AQ017410 CIT-HSP-2
     35   33.2   17.3    613  6  CD543738              CD543738 B0251E12-
     36   33.2   17.3    616  7  CF898187              CF898187 A0235E06-
     37   33.2   17.3    621  6  CD541777              CD541777 B0235A12-
     38   33.2   17.3    623  6  CA883270              CA883270 B0102B10-
     39   33.2   17.3    630  9  AQ268787              AQ268787 RPCI11-74
     40   33.2   17.3    642  2  BB658280              BB658280 BB658280
     41   33.2   17.3    657 11   CR961164              CR961164 Homo sapi
     42   33.2   17.3    663  7  CF904613              CF904613 A0424D01-
     43   33.2   17.3    771  7  CN455104              CN455104 UI-M-HN0-
     44   33.2   17.3    774  7  CO433921              CO433921 UI-M-HX0-
     45   33.2   17.3    777  7  CN459909              CN459909 UI-M-HB0-


                              ALIGNMENTS


RESULT 1
CE144436
LOCUS       CE144436                 673 bp    DNA     linear   GSS 25-SEP-2003
DEFINITION  tigr-gss-dog-17000371273067 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE144436
VERSION     CE144436.1  GI:35257693
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..673
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             21.2%;  Score 40.8;  DB 9;  Length 673;
  Best Local Similarity   56.8%;  Pred. No. 0.1;
  Matches   75;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         23 TGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGA 82
              ||||||| |   ||| || |  || | | | ||| |||||||||   |  |  |||||
```

```
Db        326 TGGCTTGTCCTTCATTATGTCCTCAGCTCCTTCTGTTTTCAATAGAGTTTGTCAAATAAG 385

Qy         83 TGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGAC 142
              |   || |   | ||||   |  |||||| ||  |  |  ||  | |  ||| |||
Db        386 TAGCAAGAGTTTGATAAGCAGGAGGTTTTAGCTTAGTCTGGCTATCGGTTCAGTGACTGT 445

Qy        143 AACATGGACTAA 154
              |   || | |||
Db        446 ATTCTGTAGTAA 457


RESULT 2
DU019646/c
LOCUS       DU019646                 826 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  276061 Tomato MboI BAC Library Lycopersicon esculentum genomic
            clone SL_MboI0087H18 5, genomic survey sequence.
ACCESSION   DU019646
VERSION     DU019646.1  GI:72421548
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 826)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 286317
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 87  row: H  column: 18
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 16
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1..826
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="SL_MboI0087H18"
                     /lab_host="E. coli"
                     /clone_lib="Tomato MboI BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: MboI"
ORIGIN

  Query Match            19.2%;  Score 36.8;  DB 10;  Length 826;
  Best Local Similarity   59.6%;  Pred. No. 1.9;
  Matches   62;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy         70 TCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAG 129
              || ||  | |  |||||  || |||| ||   | | || ||||  |  |  |  |  | |
Db        205 TCTGGCTACTCAATGTGTTAAACCTTTTATTTCTCTGGATTTCTGACTAGGAAACCTTCG 146

Qy        130 TATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCT 173
                | |||||||| || ||| | | |  || | |  |||||| ||
Db        145 ATTAATTGATGAAAATATGAAGTGAAAAATAATTATTGTCCTCT 102


RESULT 3
BH039236/c
LOCUS       BH039236                 749 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-254N24.TV RPCI-24 Mus musculus genomic clone
            RPCI-24-254N24, genomic survey sequence.
ACCESSION   BH039236
VERSION     BH039236.1  GI:14816964
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-254N24.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 254  row: N  column: 24
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..749
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-254N24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /clone_lib="RPCI-24"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
ORIGIN

  Query Match            19.0%;  Score 36.4;  DB 9;  Length 749;
  Best Local Similarity   54.5%;  Pred. No. 2.5;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy         13 GGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCG 72
              |||  |   ||||||| ||| ||| |    || |||||     | ||  || || |||
Db        456 GGCATTCTGGTGGCTTCACATGCAAATCACTGCCTGATGATAATGTTGCCAGTATCTTGC 397

Qy         73 GGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTAT 132
                | | | | |||||| ||| | ||  |       |       ||| |   ||||  |||
Db        396 ATGCATTGGGTGTGAATACCATAATGCAGTAGCCCTCCATTTTGAAGCCATCGCTGTTAT 337

Qy        133 CATTGATGACAACA 146
              |  ||||  ||| |
Db        336 CTCTGATTGCAAAA 323


RESULT 4
DN671770
LOCUS       DN671770                1275 bp    mRNA    linear   EST 29-MAR-2005
DEFINITION  CFW64-F08.x1d-t SHGC-CFW Gasterosteus aculeatus cDNA clone
            CFW64-F08 3', mRNA sequence.
ACCESSION   DN671770
VERSION     DN671770.1  GI:61991827
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1275)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Ave, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 64
            High quality sequence start: 3
            High quality sequence stop: 841.
FEATURES             Location/Qualifiers
     source          1..1275
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Conner Creek sticklebacks, WA USA"
                     /db_xref="taxon:69293"
                     /clone="CFW64-F08"
                     /sex="mixed male and female"
                     /tissue_type="gills"
                     /dev_stage="adult"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="SHGC-CFW"
                     /note="Vector: Express 1; Total and poly A+ RNA was
                     isolated from the indicated stickleback tissue, and a cDNA
                     library was constructed in the Express 1 plasmid vector by
                     Open Biosystems. First strand cDNA synthesis was primed
                     with an 54 bp linker primer containing an oligodT sequence
                     preceeded by a synthetic NotI site (first strand primer:
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following
                     second strand synthesis, cDNAs were made blunt at the end
                     corresponding to the original 5 prime end of mRNA, and
                     cloned directionally into the NotI and EcoRV sites of
                     Express 1. Note that the EcoRV site is typically destroyed
                     in the blunt end cloning, leaving a junction of the form
                     'xxxATC'(where is ATC is the second half of the EcoRV
                     site, and xxx is derived from the cDNA sequence). A map of
                     the Express 1 vector is available at:
                     http://www.openbiosystems.com/cdna_library_construction_fa
                     q.php#8 The primary library was transformed and amplified
                     in DH10B (T1 phage resistant) bacteria. Clones available
                     from Open Biosystems:
                     http://www.openbiosystems.com/stickleback"
ORIGIN

  Query Match             18.2%;  Score 35;  DB 8;  Length 1275;
  Best Local Similarity   66.7%;  Pred. No. 7.9;
  Matches   50;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         34 GCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCC 93
              ||  |  |||  ||||   ||  | ||||   ||||| |||||||  ||  | ||||| | 
Db       1087 GCCAACACTTTCCTGAAAAGTGGAATTTCGGAACCTTTGGGGAAAAGGAGGCGAAAAACA 1146

Qy         94 TTATAAAACGCGGGT 108
               | |||||| |||||
Db       1147 GTTTAAAACCCGGGT 1161


RESULT 5
BG542199
LOCUS       BG542199                 852 bp    mRNA    linear   EST 03-APR-2001
DEFINITION  602571637F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4696041 5',
            mRNA sequence.
ACCESSION   BG542199
VERSION     BG542199.1  GI:13534432
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1522  row: j  column: 10
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1..852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4696041"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             18.0%;  Score 34.6;  DB 2;  Length 852;
  Best Local Similarity   55.4%;  Pred. No. 9.7;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         38 AATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTAT 97
              | | | |||| ||| | | |      |   | |||||||  ||  ||   |  ||  |||
Db        446 ACTATGGTCTCATTGGCCAACGCCACAATACATCGGGGACCTAAGTGACCAGCCCAATAT 505

Qy         98 AAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCA 157
               ||  ||||| | |  |  | ||  |  | |||  ||||  |    ||||||||||  | 
Db        506 CAACGGCGGGCTCTATCCTATACTCGGACAAGTCACATTTTTCCATACATGGACTACCCC 565

Qy        158 A 158
              |
Db        566 A 566


RESULT 6
BH378221
LOCUS       BH378221                 590 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-119J16.TR ND-TAM Anopheles gambiae genomic clone
            AG-ND-119J16, genomic survey sequence.
ACCESSION   BH378221
VERSION     BH378221.1  GI:17324363
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
```

```
            Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
            Gardner,M.J. and Collins,F.H.
  TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
  JOURNAL   Mol. Genet. Genomics 268 (6), 720-728 (2003)
   PUBMED   12655398
COMMENT     Other_GSSs: AG-ND-119J16.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..590
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-119J16"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
ORIGIN

  Query Match             17.9%;  Score 34.4;  DB 9;  Length 590;
  Best Local Similarity   53.8%;  Pred. No. 10;
  Matches   71;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          38 AATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTAT 97
               ||| || ||| ||   ||||| ||| |   || | |    ||     | ||      | |
Db         459 AATTTTATCTTATCTTTCTATCTTCCACTTCTACCGAAGCATTAGAATTAATTTGGATGT 518

Qy          98 AAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCA 157
               || |  |  ||||| |  ||| | |  ||| || |||   |||| ||    |||  |||
Db         519 AACAGACCAGTTTTAGTTGAATCTTAAGCTCGTCTCAACAATGAAAAACGTGACGCAGCT 578

Qy         158 AAAGTGCTTGTC 169
                ||   || | |
Db         579 CAACCCCTAGCC 590


RESULT 7
AZ964976/c
LOCUS       AZ964976                 600 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0234D24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0234D24 R, genomic survey sequence.
ACCESSION   AZ964976
VERSION     AZ964976.1  GI:13836203
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
```

```
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0234  row: D  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0234D24"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 9;  Length 600;
  Best Local Similarity   56.8%;  Pred. No. 12;
  Matches   63;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy          51 TCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTT 110
               | | |||| | ||| ||||  ||  | |  | |   | || ||||||   ||    ||
Db         330 TAGCCTATCTGCAAAACCTCAGGATACAGGGGTTGAAGAAGCCTTATTTCACAACTGTCC 271

Qy         111 TCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |  || | | | ||    || ||||||   | |||| ||| ||   ||| |
Db         270 TAACATATATAAGCCTATGTCTCATTGTGCATAACAAGGAGTATATAAAGG 220


RESULT 8
BP519798
LOCUS       BP519798                 619 bp    mRNA    linear   EST 17-SEP-2003
DEFINITION  BP519798 Hydra magnipapillata cDNA library Hydra magnipapillata
            cDNA clone hmp_17065, mRNA sequence.
ACCESSION   BP519798
VERSION     BP519798.1  GI:34823167
KEYWORDS    EST.
SOURCE      Hydra magnipapillata
  ORGANISM  Hydra magnipapillata
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Hayakawa,S., Hwang,J.S., Nishimiya-Fujisawa,C., Ogura,A., Ikeo,K.,
```

```
            Fujisawa,T. and Gojobori,T.
  TITLE     Hydra EST project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jung Shan Hwang
            Center for Information Biology and DNA Data Bank of Japan
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-55-981-6847(ex.6898)
            Fax: 81-55-981-6848
            Email: jhwang@lab.nig.ac.jp, URL:http://www.cib.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Hydra magnipapillata"
                     /mol_type="mRNA"
                     /strain="105"
                     /db_xref="taxon:6085"
                     /clone="hmp_17065"
                     /tissue_type="whole body"
                     /dev_stage="adult budding stage"
                     /clone_lib="Hydra magnipapillata cDNA library"
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 3;  Length 619;
  Best Local Similarity   50.9%;  Pred. No. 12;
  Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          17 ATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGA 76
               |||||||||  |  |||  |  | | |     |     |   | |||| |   |     |
Db          30 ATGAAGTGGGCTTTCAAAGAAGAACATACTATAGAAAGCAGATGTCAAGAGTCTACAAAA 89

Qy          77 AATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATT 136
               |  ||| |  |  | ||  ||  ||  |  ||| | |  |||| |    || ||  ||  
Db          90 ATAAGAAGCAAGTATCCGGATCGAATTCCTGTTGTAGTTGAAAAAGCTCCTCGTTCCACA 149

Qy         137 GATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGA 175
                 | |  | || ||| ||  |||  | ||||| |||| |
Db         150 ATTCAAGATATTGACAAAAGAAAGTTTCTTGTTCCCTCA 188


RESULT 9
DN244174
LOCUS       DN244174                 651 bp    mRNA    linear   EST 29-JUN-2005
DEFINITION  ACAE-aaa24n15.g1 Hydra EST UCI 5 Hydra magnipapillata cDNA 5'
            similar to gb|AAO45172.1| hypothetical GABA(A) receptor-associated
            protein [Branchiostoma belcheri], mRNA sequence.
ACCESSION   DN244174
VERSION     DN244174.2  GI:68337862
KEYWORDS    EST.
SOURCE      Hydra magnipapillata
  ORGANISM  Hydra magnipapillata
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
            Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,
            Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
            Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagareishvili,R.,
            Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
  TITLE     WashU Hydra EST Project
  JOURNAL   Unpublished (2002)
COMMENT     On Mar 1, 2005 this sequence version replaced gi:60406628.
            Contact: Hans Bode
            WashU Hydra EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library material provided by Hans Bode & Dirk Lindgens, Univ. of
            Calif., Irvine Library constructed by Dirk Lindgens, Univ. of
            Calif. Irvine Library sequenced by Washington University Genome
            Sequencing Center For information on obtaining a clone please
            contact: Hans Bode (hrbode@uci.edu)
            This trace has been recalled with phred
            original value before phred recall for SL was 100
            original value before phred recall for SR was 790
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Hydra magnipapillata"
                     /mol_type="mRNA"
                     /strain="105"
                     /db_xref="taxon:6085"
                     /lab_host="DH10B"
                     /clone_lib="Hydra EST UCI 5"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I;
                     a.1st strand cDNA was primed with a Not I primer-adapter
                     (5' - pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3')
                     b.Double-stranded cDNA was ligated to Sal I adapter,
                     digested with Not I and cloned into the pSPORT 1-vector
                     pre-cut with Not I and Sal I. c.The ligation mix was
                     transformed into DH10B cells. d.The cells were grown in
                     SOC = 5g yeast, 20g tryptone, 0.5 g NaCl, 10 mM MgSO4, 10
                     mM MgCl, 0.2% glucose/Liter, (no antibiotic). e.DMSO was
                     added to a final conc. of 10% as a cryoprotectant.and
                     frozen f.The titre before freezing was determined as
                     ~2400/100 ul. Assuming a 10% loss upon freezing, the titre
                     is probably ~2100/ 100 ul. g.9 tubes each containing ~
                     2100 clones/100 ul [ = total of ~19,000] are enclosed.
                     h.The frequency of vectors containing inserts is 96% as
                     determined by digestion check after picking 24 clones,
                     miniprep and subsequent digestion with Not I and Sal I.
                     i.A low level of 32P was used in the cDNA synthesis
                     procedure. The level measured by holding a Geiger Counter
                     next to a tube was background."
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 8;  Length 651;
  Best Local Similarity   50.9%;  Pred. No. 12;
  Matches   81;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          17 ATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGA 76
               |||||||||  |  |||  |  |   | | | |     |   | |||| |   |     |
Db          81 ATGAAGTGGGCTTTCAAAGAAGAATATTTTTTAGAAAGCAGATGTCAAGAGTCTACAAAA 140

Qy          77 AATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATT 136
               |  ||| |  |  | ||  ||  ||  |  ||| | |  | | |    || ||  ||  
Db         141 ATAAGAAGCAAGTATCCGGATCGAATTCCTGTTGTAGTTGAGAAAGCTCCTCGTTCCACA 200

Qy         137 GATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGA 175
                 | |  | || ||| ||  |||  | ||||| |||| |
Db         201 ATTCAAGATATTGACAAAAGAAAGTTTCTTGTTCCCTCA 239


RESULT 10
CR281388
LOCUS       CR281388                 878 bp    mRNA    linear   EST 27-FEB-2004
DEFINITION  CR281388 Oryza sativa library (Han B) Oryza sativa cDNA clone
            y620d12p5, mRNA sequence.
ACCESSION   CR281388
VERSION     CR281388.1  GI:44667954
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Ying,K., Li,Y., Guan,J.P., Zhu,J.J.,
            Zhao,Q., Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L.,
            Weng,Q.J., Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T.,
            Zhang,Y.J., Lu,Y., Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X.,
            Zhang,L., Lan,L.F., Chen,W., Wu,S.A. and Xue,Y.B.
  TITLE     Rice cDNA EST clone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Han Bin
```

```
            National Center for Gene Research
            Chinese Academy of Sciences
            500# Cao Bao Road, Shanghai 200233, China
            Email: bhan@ncgr.ac.cn
            Clone requests: bhan@ncgr.ac.cn
             This is rice cdna est clone
            Web site: http://www.ncgr.ac.cn.
FEATURES             Location/Qualifiers
     source          1..878
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="y620d12p5"
                     /clone_lib="Oryza sativa library (Han B)"
ORIGIN

  Query Match             17.8%;  Score 34.2;  DB 7;  Length 878;
  Best Local Similarity   49.2%;  Pred. No. 13;
  Matches   90;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          2 CTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTT 61
              |||| |   |   ||||||  |  |  |    |  | ||    |  | | |  | |||
Db        526 CTGGTCATTGTACCCATGAGGGTGACTATTGACGAGAAGTCGGGTTTACTGTTGTATCAG 585

Qy         62 CAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACA 121
               | |    ||  ||| | |  | | ||||         || |||||  |||||||  ||
Db        586 GATTTGGGTCTTGGATAAAATTCTTAAAATTTGGACAAGCGCGGTATTTCGCTCAATGG 645

Qy        122 TGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAAG 181
                | || ||||| |    ||| |   |  || | || | ||    |  ||||||||   |
Db        646 CCCTCTGGTATAAAATCTGAGACACTTTACAACGCTACAGAAGGTACCCCCTGACAGCAA 705

Qy        182 AAG 184
               ||
Db        706 TAG 708


RESULT 11
CNS079HE/c
LOCUS       CNS079HE                 945 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T3 end of clone BB0AA029A02 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL435288
VERSION     AL435288.1  GI:12218701
KEYWORDS    GSS.
SOURCE      Pichia angusta
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
   PUBMED   11152888
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
```

```
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism="Pichia angusta"
                     /mol_type="genomic DNA"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA029A02"
                     /clone_lib="BB0AA"
                     /note="end : T3"
     misc_feature    <4..>944
                     /note="similar to Saccharomyces cerevisiae ORF YMR290c [
                     HAS1 ; helicase associated with Set1p ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

  Query Match             17.6%;  Score 33.8;  DB 11;  Length 945;
  Best Local Similarity   56.9%;  Pred. No. 18;
  Matches   62;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              || |||  ||||||| |||||||||  |||||  ||||  |    || |  | ||  || |
Db        387 CAGAATTCTGTCTGCTTCGTCTATAATCAATGTCTTCAAGTTCTTAAAGATAAATCCCTT 328

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACA 143
                ||    ||  ||  ||    || |   |   |||  || |||| | |
Db        327 CGTATTTTGCAAGTGATCCAGCAATCTACCTGGAGTGGCAATGATCAAA 279


RESULT 12
CNS0757U/c
LOCUS       CNS0757U                 845 bp    DNA     linear   GSS 07-JUL-2001
DEFINITION  clone BA0AB033E09 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, genomic survey sequence.
ACCESSION   AL429760
VERSION     AL429760.1  GI:12212954
KEYWORDS    GSS.
SOURCE      Kluyveromyces lactis
  ORGANISM  Kluyveromyces lactis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 845)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
   PUBMED   11152886
REFERENCE   3  (bases 1 to 845)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Kluyveromyces lactis"
                     /mol_type="genomic DNA"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB033E09"
                     /clone_lib="BA0AB"
     misc_feature    <1. .>845
                     /note="similar to Saccharomyces cerevisiae ORF YPR010c [
                     RPA135 ; DNA-directed RNA polymerase I, 135 KD subunit ]"
                     /evidence=not_experimental
ORIGIN

  Query Match             17.5%;  Score 33.6;  DB 11;  Length 845;
  Best Local Similarity   54.1%;  Pred. No. 20;
  Matches   66;  Conservative    1;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         23 TGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGA 82
              | | |||| | |  | |||| |||| |    |||| | | |||||| |   || | |||
Db        165 TAGTTTGAGATGTTTGATCTGGTCTAACGGATCTAATCTGAATACCATATTGGGAGTAGG 106

Qy         83 TGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGAC 142
               |  |   |  |||  ||| |  ||| |    | |  |||  | | || ||   :   ||
Db        105 AGGCACCTCTGTTAGCAAATGAAGGTCTAATGATAGCCATTGGATGGTTTCTACKTGCAC 46

Qy        143 AA 144
              ||
Db         45 AA 44


RESULT 13
CW393812/c
LOCUS       CW393812                 463 bp    DNA     linear   GSS 01-NOV-2004
DEFINITION  fsbb001f081a17f0 Sorghum methylation filtered library (LibID: 104)
            Sorghum bicolor genomic clone fsbb001f081a17, genomic survey
            sequence.
ACCESSION   CW393812
VERSION     CW393812.1  GI:55112256
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
```

```
            Email: jbedell@oriongenomics.com
            Plate: fsbb001f081  row: a  column: 17
            Seq primer: f Forward
            Class: methylation filtered
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="fsbb001f081a17"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN

  Query Match             17.4%;  Score 33.4;  DB 10;  Length 463;
  Best Local Similarity   57.0%;  Pred. No. 20;
  Matches   61;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         62 CAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACA 121
              |||  | ||    | ||  | || || || |||   ||  |   ||  || || ||||
Db        354 CAAGGCATTGTAAGTAAGTGCTGCGATAAGCCTCCAAAGGCCTTGGCCTTTGCTGAAAGC 295

Qy        122 TGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
               ||  | || |||  | ||||||  || || |||  ||  ||  |||
Db        294 GGCCATGGTTTCAAAGGTGACAAGGTGAACCAAGGCAAGCTGGGTGT 248


RESULT 14
CA394392
LOCUS       CA394392                 494 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  cs51a07.y1 Human Retinal pigment epithelium/choroid cDNA
            (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs51a07
            5', mRNA sequence.
ACCESSION   CA394392
VERSION     CA394392.1  GI:24728898
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
            Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
  TITLE     Expressed sequence tag analysis of human RPE/choroid for the
            NEIBank Project: Over 6000 non-redundant transcripts, novel genes
            and splice variants
  JOURNAL   Mol. Vis. 8 (4), 205-220 (2002)
   PUBMED   12107410
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 51  row: a  column: 07
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="cs51a07"
                     /tissue_type="RPE/choroid"
```

```
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Retinal pigment epithelium/choroid cDNA
                     (Un-normalized, unamplified): cs"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
                     eyes (75-80 years old) yielded approximately 600 mg of
                     dissected RPE/choroid tissue.  This in turn yielded 340 ug
                     of total RNA and 7 ug of mRNA. A directionally cloned cDNA
                     library in the pCMVSPORT6 vector was constructed at Life
                     Technologies (Rockville, MD; now part of Invitrogen Corp),
                     essentially following the protocols of the SuperScript
                     Plasmid System (Invitrogen Corp.
                     <http://www.invitrogen.com/>). The library code
                     designation was cs. For this library, cDNA inserts were
                     cloned into the NotI/MluI sites of the vector. EST
                     analysis was performed on the unamplified library at the
                     NIH Intramural Sequencing Center (NISC)."
ORIGIN

  Query Match             17.4%;  Score 33.4;  DB 6;  Length 494;
  Best Local Similarity   52.5%;  Pred. No. 21;
  Matches   73;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          42 TTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAA 101
               || |||  || | ||  | | |  | || |  | ||||  || ||   | |   || ||
Db           7 TTTTCTACTTTGCCTCCTCTTACAAGCTACAAGCAAATGAATCTGCTCAGCAGAATGAAG 66

Qy         102 CGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               |     |||  ||| ||| ||    |   || ||| |     | ||   |||||||  |
Db          67 CCTATTTTTCAGCATAAATATTGCTTGCCATGATTAAAATAGAAATCAGCTAAGCATGAT 126

Qy         162 TGCTTGTCCCCTGACCCAA 180
               | |||  |   ||  ||||
Db         127 TTCTTTACTGGTGGACCAA 145


RESULT 15
BH885133
LOCUS       BH885133                 728 bp    DNA     linear   GSS 05-AUG-2002
DEFINITION  hw61c11.b1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone hw61c11 5', genomic survey sequence.
ACCESSION   BH885133
VERSION     BH885133.1  GI:22121030
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: hw61  row: c  column: 11
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 728.
FEATURES             Location/Qualifiers
     source          1. .728
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="hw61c11"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
ORIGIN

  Query Match             17.4%;  Score 33.4;  DB 9;  Length 728;
  Best Local Similarity   62.7%;  Pred. No. 23;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy         110 TTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
               |||  ||| ||||||| || || ||  ||| |||| | | | |||   |  || ||  ||
Db         171 TTCATAGAGACATGCGATAATAGCAGGGATAACAATAAGAAATAATTCATCGTCCTCTTC 230

Qy         170 CCCTGACCCAAGAAGGATGCTTT 192
               | |     |||| ||||  || |
Db         231 CTCCTCTTCAAGCAGGAACCTAT 253


Search completed: March  7, 2006, 13:19:34
Job time : 2366.02 secs
```

```
                    GenCore version 5.1.7
           Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 11:56:56 ; Search time 134.579 Seconds
                                           (without alignments)
                                           2535.988 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:      2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/1_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/H_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                7:  /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
                8:  /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
                9:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                     Description
------------------------------------------------------------------------
    1    32.6    17.0    278  3  US-09-513-999C-31601     Sequence 31601, A
c   2    31.2    16.2  22404  3  US-09-949-016-11765      Sequence 11765, A
c   3    31.2    16.2  22404  3  US-09-949-016-15398      Sequence 15398, A
c   4    31.2    16.2  26167  3  US-09-454-466-1          Sequence 1, Appli
c   5    31.2    16.2  46745  3  US-09-949-016-13964      Sequence 13964, A
    6    30.6    15.9    601  3  US-09-949-016-48509      Sequence 48509, A
c   7    30.6    15.9  58768  3  US-09-949-016-13175      Sequence 13175, A
    8    30.4    15.8   1569  3  US-09-710-279-603        Sequence 603, App
    9    30.4    15.8   3204  3  US-09-710-279-3485       Sequence 3485, Ap
c  10    29.8    15.5  10303  3  US-09-634-238-410        Sequence 410, App
c  11    29.8    15.5  99797  3  US-09-949-016-15255      Sequence 15255, A
c  12    29.6    15.4    601  3  US-09-949-016-187277     Sequence 187277,
c  13    29.6    15.4  36855  3  US-09-949-016-17095      Sequence 17095, A
   14    29.4    15.3    768  3  US-09-134-000C-3331      Sequence 3331, Ap
c  15    29.2    15.2    601  3  US-09-949-016-142351     Sequence 142351,
   16    29.2    15.2   2385  3  US-09-949-016-4010       Sequence 4010, Ap
   17    29.2    15.2   3077  3  US-09-461-912A-27        Sequence 27, Appl
   18    29.2    15.2   3092  2  US-08-426-627-3          Sequence 3, Appli
   19    29.2    15.2   3126  2  US-08-477-396A-3         Sequence 3, Appli
   20    29.2    15.2   3253  2  US-08-426-627-5          Sequence 5, Appli
   21    29.2    15.2   4623  3  US-09-949-016-15769      Sequence 15769, A
   22    29.2    15.2 160552  3  US-09-593-828-11         Sequence 11, Appl
   23      29    15.1    601  3  US-09-949-016-44696      Sequence 44696, A
   24      29    15.1    832  3  US-09-621-976-2813       Sequence 2813, Ap
   25      29    15.1  36791  3  US-09-949-016-16390      Sequence 16390, A
   26      29    15.1  36791  3  US-09-949-016-16391      Sequence 16391, A
   27      29    15.1  42479  3  US-09-949-016-16631      Sequence 16631, A
   28      29    15.1  42479  3  US-09-949-016-16632      Sequence 16632, A
c  29      29    15.1  48940  3  US-09-949-016-16402      Sequence 16402, A
c  30      29    15.1  91062  3  US-09-949-016-13019      Sequence 13019, A
   31      29    15.1  92407  3  US-09-596-002-36         Sequence 36, Appl
c  32      29    15.1 1830121  3  US-09-557-884-1          Sequence 1, Appli
c  33      29    15.1 1830121  3  US-09-643-990A-1         Sequence 1, Appli
c  34      29    15.1 1830121  3  US-10-158-865-1          Sequence 1, Appli
   35    28.8    15.0 636591  3  US-09-949-016-11808      Sequence 11808, A
   36    28.8    15.0 636591  3  US-09-949-016-13388      Sequence 13388, A
   37    28.6    14.9 189560  3  US-09-949-016-17202      Sequence 17202, A
c  38    28.6    14.9 263693  3  US-09-949-016-12386      Sequence 12386, A
c  39    28.6    14.9 263694  3  US-09-949-016-16915      Sequence 16915, A
   40    28.4    14.8  50381  3  US-09-949-016-17122      Sequence 17122, A
c  41    28.4    14.8  78125  3  US-09-949-016-16006      Sequence 16006, A
c  42    28.2    14.7  18264  3  US-09-949-016-16469      Sequence 16469, A
   43      28    14.6  61178  3  US-09-949-016-17369      Sequence 17369, A
   44      28    14.6 194537  3  US-09-949-016-12928      Sequence 12928, A
   45      28    14.6 201529  3  US-09-949-016-12740      Sequence 12740, A

                              ALIGNMENTS


RESULT 1
US-09-513-999C-31601
; Sequence 31601, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 31601
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 152
;   OTHER INFORMATION: k=g or t
US-09-513-999C-31601

  Query Match             17.0%;  Score 32.6;  DB 3;  Length 278;
  Best Local Similarity   51.0%;  Pred. No. 0.22;
  Matches   77;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         18 TGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAA 77
              | | |||| |||       |  |   || | |     || | |||||||  || |||||
Db         99 TCATGTGGATTGTTTTCTGTCTTTAGGTATTAAAATGCTGTATTCAATAATTTKGGGGAC 158

Qy         78 ATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTG 137
              ||| |||| |   | | |||| |  |  |  | ||   |  |||| |   |    || ||
Db        159 ATATATGTCATTTCACATATATACAGATGTGTATCAGGGTGACATTCCAAAAGTGCACTG 218

Qy        138 ATGACAACATGGACTAAGCAAAAGTGCTTGT 168
               |||      ||  || | |   ||  || |
Db        219 CTGAGTTAGAGGTATATGTATTTGTAATTTT 249


RESULT 2
US-09-949-016-11765/c
; Sequence 11765, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11765
;   LENGTH: 22404
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(22404)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11765

  Query Match            16.2%;  Score 31.2;  DB 3;  Length 22404;
  Best Local Similarity   60.7%;  Pred. No. 3.8;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      19171 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 19112

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      19111 AACACCCATGAGAACAGGGACTCA 19088

RESULT 3
US-09-949-016-15398/c
; Sequence 15398, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15398
;   LENGTH: 22404
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(22404)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15398

  Query Match            16.2%;  Score 31.2;  DB 3;  Length 22404;
  Best Local Similarity   60.7%;  Pred. No. 3.8;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      19171 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 19112

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      19111 AACACCCATGAGAACAGGGACTCA 19088

RESULT 4
US-09-454-466-1/c
; Sequence 1, Application US/09454466
; Patent No. 6858425
; GENERAL INFORMATION:
;  APPLICANT: Heus, Joris Jan
;  APPLICANT: Pharming Intellectual Property B.V.
;  TITLE OF INVENTION: HUMAN ACID ALPHA GLUCOSIDASE GENE AND BOVINE ALPHA-S1
;  TITLE OF INVENTION: CASEIN GENE SEQUENCES
;  FILE REFERENCE: 016994-013720US
;  CURRENT APPLICATION NUMBER: US/09/454,466
;  CURRENT FILING DATE: 1999-12-03
;  EARLIER APPLICATION NUMBER: 60/110,859
;  EARLIER FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: 60/122,550
;  EARLIER FILING DATE: 1999-03-02
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 26167
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-454-466-1

  Query Match            16.2%;  Score 31.2;  DB 3;  Length 26167;
  Best Local Similarity   60.7%;  Pred. No. 4.1;
  Matches   51;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         71 CGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGT 130
              || | | | |||  ||||||  |  | ||   |  ||| |  ||||| |  || |  ||
Db      17958 CGAGCAGAGAGAGATGAAAAGGCAGAGAATGTGGAGGTGTCAGCAGACAGCTGGGGCAGG 17899

Qy        131 ATCATTGATGACAACATGGACTAA 154
              | ||   |||| |||| ||||| |
Db      17898 AACACCCATGAGAACAGGGACTCA 17875


RESULT 5
US-09-949-016-13964/c
; Sequence 13964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13964
;   LENGTH: 46745
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(46745)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13964
```

```
  Query Match             16.2%;  Score 31.2;  DB 3;  Length 46745;
  Best Local Similarity   54.3%;  Pred. No. 5.1;
  Matches   63;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          46 CTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCG 105
               ||| |||  ||  | |||  ||||| || |  ||| | ||    ||| || ||      |
Db       40974 CTGCTTCCACTGCTGTCATGACCTTAGGTGGCATAAAAGTCCTTACCTTTCTACCCTAAG 40915

Qy         106 GGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAG 161
               | |   |  |    | |    ||||||| || |  ||| |   |  ||||  ||||
Db       40914 GATGGGCCAATTCTCTTCTTTTAGTATCTTTCAGAACAGCTGTGCTTAAGTTAAAG 40859


RESULT 6
US-09-949-016-48509
; Sequence 48509, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48509
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-48509

  Query Match             15.9%;  Score 30.6;  DB 3;  Length 601;
  Best Local Similarity   58.1%;  Pred. No. 1.5;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
               || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db         149 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 208

Qy         116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
                ||| ||| |||   || ||  ||||  | |||
Db         209 CAAAGATGAGCTGAAATAATACATGAATAAATG 241


RESULT 7
US-09-949-016-13175/c
; Sequence 13175, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13175
;   LENGTH: 58768
```

```
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(58768)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13175

  Query Match             15.9%;  Score 30.6;  DB 3;  Length 58768;
  Best Local Similarity   58.1%;  Pred. No. 9.1;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
               || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db       43027 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 42968

Qy         116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
                ||| ||| |||   || ||  ||||  | |||
Db       42967 CAAAGATGAGCTGAAATAATACATGAATAAATG 42935


RESULT 8
US-09-710-279-603
; Sequence 603, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 603
;   LENGTH: 1569
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-603

  Query Match             15.8%;  Score 30.4;  DB 3;  Length 1569;
  Best Local Similarity   57.3%;  Pred. No. 2.5;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
               ||  |  |||   ||| |   |  | ||   | || |   |   ||||||||||    ||
Db          87 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 146

Qy         109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
               |  ||| || | ||  | |||| | ||| |||| ||
Db         147 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 182


RESULT 9
US-09-710-279-3485
; Sequence 3485, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3485
;   LENGTH: 3204
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3485

  Query Match             15.8%;  Score 30.4;  DB 3;  Length 3204;
  Best Local Similarity   57.3%;  Pred. No. 3.4;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         49 ATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGT 108
              ||  |  |||   ||| |   |  | ||   | || |   |    |||||||||    ||
Db        542 ATATGAATATGAACAAAATAATGAGCGATCCATATCTCTCACTGCTTATAAAACTAATGT 601

Qy        109 TTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
              |  ||| || | ||  | |||| | ||| |||| ||
Db        602 TAACGCGGATATATTTGATAGTTTGATTAATGAAAA 637


RESULT 10
US-09-634-238-410/c
; Sequence 410, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
;  APPLICANT:  Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka J.
;  APPLICANT:  Bloksberg, Leonard, N.
;  APPLICANT:  Lubbers, Mark W.
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna C.
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul W.
;  APPLICANT:  Reid, Julian R.
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION:  Polynucleotides, materials incorporating
;  TITLE OF INVENTION:  them and methods for using them.
;  FILE REFERENCE:  11000.1043U1
;  CURRENT APPLICATION NUMBER: US/09/634,238
;  CURRENT FILING DATE:  2000-08-08
;  NUMBER OF SEQ ID NOS:  422
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 410
;   LENGTH: 10303
;   TYPE: DNA
;   ORGANISM: Lactobacillus rhamnosus
US-09-634-238-410

  Query Match             15.5%;  Score 29.8;  DB 3;  Length 10303;
  Best Local Similarity   60.5%;  Pred. No. 8.6;
  Matches   49;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        112 CGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCC 171
              || |||||   ||      ||  |||| ||   |||| |  | |||||||   ||||| |
Db       8088 CGAAGAAAGCAGCAAGGACATTGTTGAAGAAGTCATGCAGCAGGCAAAAGAAATTGTCGC 8029

Qy        172 CTGACCCAAGAAGGATGCTTT 192
              | |  | || | ||  |||||
Db       8028 CGGCTCGAAAATGGCGGCTTT 8008


RESULT 11
US-09-949-016-15255/c
; Sequence 15255, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15255
;   LENGTH: 99797
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99797)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15255

  Query Match             15.5%;  Score 29.8;  DB 3;  Length 99797;
  Best Local Similarity   51.1%;  Pred. No. 21;
  Matches   70;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         40 TCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAA 99
              | | |  ||| || |  ||||       ||   |  |||  |||||| |     ||| |
Db      32501 TGTAGGATGAATCCTGCATTTGATGAGACTAGTGTTAAAAGGATGTGCAGGGGGTTAGAC 32442

Qy        100 AACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAA 159
                 |||||     |||| |||| |  |   || ||      | ||||        | ||
Db      32441 GGGGCGGGGGAGAGCAGTAACAGGGACAGTTACCAAGCCATAAAACAAACGAGCTGAAAC 32382

Qy        160 AGTGCTTGTCCCCTGAC 176
              | | |||| || |||||
Db      32381 AATTCTTGCCCTCTGAC 32365


RESULT 12
US-09-949-016-187277/c
; Sequence 187277, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187277
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-187277

  Query Match             15.4%;  Score 29.6;  DB 3;  Length 601;
  Best Local Similarity   59.5%;  Pred. No. 3.3;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         96 ATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAG 155
              |||  || |  ||  ||   ||||  | |     |||| || ||||||| |||  | |||
Db        108 ATATGACCCCAGTAGTCTTTGAAAGCTCCTTCCCTATCTTTTATGACAAGATGTTCCAAG 49

Qy        156 CAAAAGTGCTTGTCCCCTGACCCA 179
              |  |  ||  | |  |||| ||||
Db         48 CTTATTTGTATATTTCCTGCCCCA 25


RESULT 13
```

```
US-09-949-016-17095/c
; Sequence 17095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17095
;   LENGTH: 36855
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17095

  Query Match             15.4%;  Score 29.6;  DB 3;  Length 36855;
  Best Local Similarity   59.5%;  Pred. No. 17;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         96 ATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAG 155
              |||  || |  ||  ||   ||||  | |     |||| || ||||||| |||  | |||
Db      29826 ATATGACCCCAGTAGTCTTTGAAAGCTCCTTCCCTATCTTTTATGACAAGATGTTCCAAG 29767

Qy        156 CAAAAGTGCTTGTCCCCTGACCCA 179
              |  |  ||  | |  |||| ||||
Db      29766 CTTATTTGTATATTTCCTGCCCCA 29743


RESULT 14
US-09-134-000C-3331
; Sequence 3331, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3331
;   LENGTH: 768
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-3331

  Query Match             15.3%;  Score 29.4;  DB 3;  Length 768;
  Best Local Similarity   63.4%;  Pred. No. 4.2;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         97 TAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGC 156
              | |  ||  ||||  |||| ||||| | |  | || ||||||| ||   |   |  ||
Db        327 TTATGCGACGGTTGGCGCAAAAACAAGTGTCAATAACATTGATTACTTTAAAAATGAACT 386

Qy        157 AAAAGTGCTTG 167
              ||||| |||||
Db        387 AAAAGAGCTTG 397


RESULT 15
US-09-949-016-142351/c
; Sequence 142351, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 142351
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-142351

  Query Match             15.2%;  Score 29.2;  DB 3;  Length 601;
  Best Local Similarity   54.7%;  Pred. No. 4.5;
  Matches   58;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy         55 CTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGC 114
              ||| | | || |   ||    ||| |||  |||  |    ||||||| | ||||| || |
Db        265 CTAGTGTTAAAAAACTCACACAAAAAGAGCTGAGGAAGGCTATAAAAAGAGGGTTCTCAC 206

Qy        115 AGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAA 160
              |     ||||   |  |  |    |  ||  |  ||||  ||||||
Db        205 ACTTGGATGCCTAATAACAAAAACTATCACAAAAGACTCTGCAAAA 160


Search completed: March  7, 2006, 14:11:45
Job time : 139.579 secs
```

```
                  GenCore version 5.1.7
         Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 12:14:00 ; Search time 526.953 Seconds
                                           (without alignments)
                                           3013.020 Million cell updates/sec

Title:          US-10-782-899-1_COPY_1_192
Perfect score:  192
Sequence:       1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:      19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_Main:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                   %
Result             Query
  No.    Score     Match Length DB  ID                       Description
--------------------------------------------------------------------------------
    1      192   100.0    600  3  US-09-816-391A-1          Sequence 1, Appli
    2      192   100.0    600  8  US-10-782-899-1           Sequence 1, Appli
c   3    175.2    91.2 2256646  7  US-10-470-565-1           Sequence 1, Appli
c   4     34.6    18.0    425  8  US-10-425-115-60228       Sequence 60228, A
c   5     33.2    17.3    824  5  US-10-027-632-167810      Sequence 167810,
c   6     33.2    17.3    824  6  US-10-027-632-167810      Sequence 167810,
    7     32.4    16.9    484  5  US-10-027-632-6585        Sequence 6585, Ap
    8     32.4    16.9    484  6  US-10-027-632-6585        Sequence 6585, Ap
c   9     32.4    16.9   2717 10  US-11-097-143-30178        Sequence 30178, A
c  10       32    16.7 495269  7  US-10-398-221-8           Sequence 8, Appli
   11       32    16.7 3011208  7  US-10-398-221-2058        Sequence 2058, Ap
c  12     31.6    16.5    732  8  US-10-425-115-71957       Sequence 71957, A
c  13     31.6    16.5   1437  4  US-09-925-065A-691107     Sequence 691107,
c  14     31.6    16.5   1437  4  US-09-925-065A-691108     Sequence 691108,
c  15     31.2    16.2    360  8  US-10-674-124A-12097      Sequence 12097, A
   16     31.2    16.2   1485  4  US-09-925-065A-49312      Sequence 49312, A
   17     31.2    16.2   1485  4  US-09-925-065A-49313      Sequence 49313, A
c  18     31.2    16.2  18272  3  US-09-764-891-8869        Sequence 8869, Ap
   19       31    16.1   1771  4  US-09-925-065A-65588      Sequence 65588, A
c  20     30.8    16.0    639  4  US-09-925-065A-170277     Sequence 170277,
c  21     30.8    16.0    639  4  US-09-925-065A-170278     Sequence 170278,
   22     30.6    15.9    466  7  US-10-021-323-5754        Sequence 5754, Ap
c  23     30.6    15.9    549  7  US-10-260-238-5041        Sequence 5041, Ap
c  24     30.6    15.9    572  4  US-09-925-065A-308556     Sequence 308556,
c  25     30.6    15.9    572  4  US-09-925-065A-308557     Sequence 308557,
c  26     30.6    15.9    601  4  US-09-925-065A-396377     Sequence 396377,
   27     30.6    15.9    606  4  US-09-925-065A-599874     Sequence 599874,
   28     30.6    15.9    670  4  US-09-925-065A-905740     Sequence 905740,
c  29     30.6    15.9    670  4  US-09-925-065A-912520     Sequence 912520,
c  30     30.6    15.9   1099  5  US-10-027-632-10536       Sequence 10536, A
c  31     30.6    15.9   1099  6  US-10-027-632-10536       Sequence 10536, A
   32     30.4    15.8    528  4  US-09-925-065A-149916     Sequence 149916,
   33     30.4    15.8    583  8  US-10-357-930-51874       Sequence 51874, A
   34     30.2    15.7    515  3  US-09-864-761-16206       Sequence 16206, A
   35     30.2    15.7    583  4  US-09-925-065A-460994     Sequence 460994,
   36     30.2    15.7 370469  5  US-10-087-192-250         Sequence 250, App
c  37       30    15.6    201  8  US-10-719-993-41046       Sequence 41046, A
c  38       30    15.6    639  5  US-10-027-632-185956      Sequence 185956,
c  39       30    15.6    639  6  US-10-027-632-185956      Sequence 185956,
   40       30    15.6   1296  4  US-09-925-065A-684507     Sequence 684507,
   41       30    15.6   1296  4  US-09-925-065A-684508     Sequence 684508,
c  42       30    15.6   1332  6  US-10-369-493-25530       Sequence 25530, A
   43       30    15.6   7017  6  US-10-074-024-436         Sequence 436, App
   44       30    15.6  25519  5  US-10-087-192-1207        Sequence 1207, Ap
c  45       30    15.6 303172  8  US-10-719-993-6890        Sequence 6890, Ap


                                ALIGNMENTS


RESULT 1
US-09-816-391A-1
; Sequence 1, Application US/09816391A
; Patent No. US20020054865A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, Minoru
;  APPLICANT:  TANIGUCHI, Shunichiro
;  APPLICANT:  AMANO, Jun
;  APPLICANT:  YAZAWA, Kazuyuki
;  APPLICANT:  KANO, Yasunobu
;  APPLICANT:  NAKAMURA, Toshiyuki
;  APPLICANT:  SASAKI, Takayuki
;  TITLE OF INVENTION: Anaerobic bacterium as a drug for cancer gene therapy
;  FILE REFERENCE: 2001-WMC/01736
;  CURRENT APPLICATION NUMBER: US/09/816,391A
;  CURRENT FILING DATE:  2001-03-26
;  PRIOR APPLICATION NUMBER: JP 00/287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-09-816-391A-1

  Query Match             100.0%;  Score 192;  DB 3;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-54;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy        181 GAAGGATGCTTT 192
              ||||||||||||
```

```
Db          181 GAAGGATGCTTT 192


RESULT 2
US-10-782-899-1
; Sequence 1, Application US/10782899
; Publication No. US20050025745A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, MINORU
;  APPLICANT:  TANIGUCHI, SHUNICHIRO
;  APPLICANT:  AMANO, JUN
;  APPLICANT:  YAZAWA, KAZUYUKI
;  APPLICANT:  KANO, YASUNOBU
;  APPLICANT:  NAKAMURA, TOSHIYUKI
;  APPLICANT:  SASAKI, TAKAYUKI
;  TITLE OF INVENTION: ANAEROBIC BACTERIUM AS A DRUG FOR CANCER GENE THERAPY
;  FILE REFERENCE: 671308-2001.1
;  CURRENT APPLICATION NUMBER: US/10/782,899
;  CURRENT FILING DATE:  2004-02-23
;  PRIOR APPLICATION NUMBER: 09/816,391
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: JP 2000-287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-10-782-899-1

  Query Match             100.0%;  Score 192;  DB 8;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-54;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60

Qy           61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 TCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAAC 120

Qy          121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 ATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAA 180

Qy          181 GAAGGATGCTTT 192
                ||||||||||||
Db          181 GAAGGATGCTTT 192


RESULT 3
US-10-470-565-1/c
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
;  APPLICANT: Societe des Produits Nestle S.A.
;  TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
;  FILE REFERENCE: 80290/WO
;  CURRENT APPLICATION NUMBER: US/10/470,565
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: EP 01102050.0
;  PRIOR FILING DATE: 2001-01-30
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2256646
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
US-10-470-565-1

  Query Match             91.2%;  Score 175.2;  DB 7;  Length 2256646;
  Best Local Similarity   97.4%;  Pred. No. 1.2e-46;
  Matches  189;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy            1 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2242172 GCTGGGCGCGGCGGCCATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTT 2242113

Qy           61 TC--AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 118
                ||  |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2242112 TCATACCCCCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAA 2242053

Qy          119 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 178
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2242052 ACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCC 2241993

Qy          179 AAGAAGGATGCTTT 192
                ||||||||||||||
Db      2241992 AAGAAGGATGCTTT 2241979


RESULT 4
US-10-425-115-60228/c
; Sequence 60228, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 60228
;   LENGTH: 425
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_15492C.1
US-10-425-115-60228

  Query Match             18.0%;  Score 34.6;  DB 8;  Length 425;
  Best Local Similarity   51.6%;  Pred. No. 0.6;
  Matches   79;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy           16 CATGAAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGG 75
                | |||| |    ||||||  ||||   |   |  |||  | |     |||||  |   |
Db          187 CTTGAAATCTACTGACAAAGATAAGACTAAATTTTTCACCCAAAACAAATACTATAATGC 128

Qy           76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
                ||  | |      ||  | | | ||| ||   ||| |  |  |  | || ||  ||   |
Db          127 AATGACAGCAATTAATACATTTGAAATGCCTATTTACATAATACAAAGCACTGCTAGACT 68

Qy          136 TGATGACAACATGGACTAAGCAAAAGTGCTTGT 168
                |  || |||||    |||| || ||| | | ||
Db           67 TACTGCCAACAATAGCTAAACAGAAGAGGTGGT 35


RESULT 5
US-10-027-632-167810/c
; Sequence 167810, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
```

```
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167810
;   LENGTH: 824
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-167810

  Query Match             17.3%;  Score 33.2;  DB 5;  Length 824;
  Best Local Similarity   59.6%;  Pred. No. 2.5;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||| |||||||   | |  | ||    |||   ||||   ||  |  | |||
Db        384 ATATAAATGAGAAAACCAAAAGAGTAAGCAATATTTAAAAGAAGGTTGAACATGGATCTA 325

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
              |||||| |||| |   |||||||  ||  |  ||
Db        324 TGATGAGAACAAGTCATAAGCAAGGGTTGTAATC 291


RESULT 6
US-10-027-632-167810/c
; Sequence 167810, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 167810
;   LENGTH: 824
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-167810

  Query Match             17.3%;  Score 33.2;  DB 6;  Length 824;
  Best Local Similarity   59.6%;  Pred. No. 2.5;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy         76 AAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCAT 135
              | ||| ||| |||||||   | |  | ||    |||   ||||   ||  |  | |||
Db        384 ATATAAATGAGAAAACCAAAAGAGTAAGCAATATTTAAAAGAAGGTTGAACATGGATCTA 325

Qy        136 TGATGACAACATGGACTAAGCAAAAGTGCTTGTC 169
              |||||| |||| |   |||||||  ||  |  ||
Db        324 TGATGAGAACAAGTCATAAGCAAGGGTTGTAATC 291


RESULT 7
US-10-027-632-6585
; Sequence 6585, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6585
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-6585

  Query Match             16.9%;  Score 32.4;  DB 5;  Length 484;
  Best Local Similarity   60.7%;  Pred. No. 3.6;
  Matches   51;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         36 ATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTT 95
              | ||| || | |    |  ||| | | ||| ||| |||| :|||||   ||| ||   |
Db        130 AGAATATTCTTTTTCACACCTACTGTTAATCCCTCCGGGTRAATAGGCATGACAATAATA 189

Qy         96 ATAAAACGCGGGTTTTCGCAGAAA 119
              || |||   |   |||||  |||||
Db        190 ATGAAATACTTTTTTTCTGAGAAA 213


RESULT 8
US-10-027-632-6585
; Sequence 6585, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
```

```
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6585
;   LENGTH: 484
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-6585

  Query Match             16.9%;  Score 32.4;  DB 6;  Length 484;
  Best Local Similarity   60.7%;  Pred. No. 3.6;
  Matches   51;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy            36 ATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTT 95
                 | ||| || | |    |  ||| | | ||| ||| |||| :|||||   ||| ||   |
Db           130 AGAATATTCTTTTCACACCTACTGTTAATCCCTCCGGGTRAATAGGCATGACAATAATA 189

Qy            96 ATAAAACGCGGGTTTTCGCAGAAA 119
                 || |||  |   |||||  |||||
Db           190 ATGAAATACTTTTTTTCTGAGAAA 213


RESULT 9
US-11-097-143-30178/c
; Sequence 30178, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30178
;   LENGTH: 2717
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-30178

  Query Match             16.9%;  Score 32.4;  DB 10;  Length 2717;
  Best Local Similarity   56.6%;  Pred. No. 8.2;
  Matches   60;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy            39 ATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATA 98
                 ||  |  | ||| | |  |  | |||||| |      |  || || |  | |  |||| |
Db           369 ATAGTTACAGATACATAAAAATACAATACATAGCAAAAGGTACATATATAGAGACTTAAA 310
```

```
Qy            99 AAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAA 144
                 | ||||  |||||| | |||  || |  ||  ||  ||   |||||
Db           309 ATACGCTTGTTTTCACTGAATTATTCAATACGATTTTAGAGACAA 264

RESULT 10
US-10-398-221-8/c
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 495269
;   TYPE: DNA
;   ORGANISM: Listeria innocua
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(end)
;   OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

  Query Match             16.7%;  Score 32;  DB 7;  Length 495269;
  Best Local Similarity   50.7%;  Pred. No. 1.4e+02;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy            35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
                 ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db        229177 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 229118

Qy            95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
                  || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db        229117 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 229058

Qy           155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
                 ||    ||    |   ||  | | ||  | ||
Db        229057 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 229026


RESULT 11
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
;  APPLICANT: KUNST, Frederik
;  APPLICANT: GLASER, Philippe
;  TITLE OF INVENTION: Listeria innocua, genome and applications
;  FILE REFERENCE: 344 702 - US
;  CURRENT APPLICATION NUMBER: US/10/398,221
;  CURRENT FILING DATE:  2003-03-27
;  PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;  PRIOR FILING DATE: 2001-10-04
;  PRIOR APPLICATION NUMBER: FR 00/12 697
;  PRIOR FILING DATE: 2000-10-04
;  NUMBER OF SEQ ID NOS: 4025
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
;   LENGTH: 3011208
;   TYPE: DNA
;   ORGANISM: Listeria innocua
US-10-398-221-2058

  Query Match             16.7%;  Score 32;  DB 7;  Length 3011208;
```

```
  Best Local Similarity   50.7%;  Pred. No. 3.1e+02;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              ||| | || |       | || ||| |||     |||||| | |   || ||||||| ||
Db    1529778 CATGAACTCGGACACGCCTTCCATTCTCATGTTATTCGGGAAGAGCCATTTGAAAACACT 1529837

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               || |  |   |  | | || |||||     ||| | |   |||    |  || | | |
Db    1529838 GATTACGCAATGAATGTTGCGGAAACTGCTTCTACTTTTGCTGAAATGATTATTGCCGAT 1529897

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAAGGA 186
              ||    ||    |   ||  | | ||  | ||
Db    1529898 GCTTCTGTAAAAGATGCCAAAACAAAAGAAGA 1529929


RESULT 12
US-10-425-115-71957/c
; Sequence 71957, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 71957
;   LENGTH: 732
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(732)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_165627C.1
US-10-425-115-71957

  Query Match             16.5%;  Score 31.6;  DB 8;  Length 732;
  Best Local Similarity   58.5%;  Pred. No. 8.1;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         54 TCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCG 113
              ||| || | |||   |      | ||||| ||||| |  ||||| ||| |  | |  | |
Db        361 TCTTTTGTTAATTTGGTGCAACATATAGAGGTGAACAAGCTTATTAAAGGAAGATCCTTG 302

Qy        114 CAGAAACATGCGCTAGTATCATTGATGACAACAT 147
               | || || | | |  ||  || ||||| |  ||
Db        301 TACAATCAAGAGTTCATACAATAGATGAAACGAT 268


RESULT 13
US-09-925-065A-691107/c
; Sequence 691107, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 691107
;   LENGTH: 1437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
;   LOCATION: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
;   LOCATION: 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,
;   LOCATION: 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
;   LOCATION: 61, 62
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
;   LOCATION: 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
;   LOCATION: 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102
;   OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691107

  Query Match             16.5%;  Score 31.6;  DB 4;  Length 1437;
  Best Local Similarity   55.5%;  Pred. No. 11;
  Matches   61;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         53 GTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTC 112
              |||| | |  ||  |    | |  |||| |||||||| |  ||    |||    || ||
Db        479 GTCTTTGTGAAAGCCAAATGTGATAATACATGTGAAAGCATTTTGTGAACTAAAGTGTTT 420

Qy        113 GCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
              | | |||||  |   |    |||  ||| |   |||||  ||| | | ||
Db        419 GTACAAACAACCTAAATGTCCATCAATGGCTGAATGGATAAAGAAGAGGT 370


RESULT 14
US-09-925-065A-691108/c
; Sequence 691108, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 691108
;   LENGTH: 1437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
;   LOCATION: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
;   LOCATION: 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,
;   LOCATION: 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
;   LOCATION: 61, 62
```

```
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
;   LOCATION: 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
;   LOCATION: 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102
;   OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691108

  Query Match             16.5%;  Score 31.6;  DB 4;  Length 1437;
  Best Local Similarity   55.5%;  Pred. No. 11;
  Matches   61;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         53 GTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTC 112
              |||| | |  ||  |    | |  |||| |||||||| |  ||    |||    || ||
Db        479 GTCTTTGTGAAAGCCAAATGTGATAATACATGTGAAAGCATTTTGTGAACTAAAGTGTTT 420

Qy        113 GCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
              | | |||||  |   |    |||  ||| |   |||||  ||| | | ||
Db        419 GTACAAACAACCTAAATGTCCATCAATGGCTGAATGGATAAAGAAGAGGT 370


RESULT 15
US-10-674-124A-12097/c
; Sequence 12097, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 12097
;   LENGTH: 360
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: AC007735.2_88547
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 7
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 54730897
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 7491
US-10-674-124A-12097

  Query Match             16.2%;  Score 31.2;  DB 8;  Length 360;
  Best Local Similarity   54.3%;  Pred. No. 7.9;
  Matches   63;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy         47 TGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGG 106
              || |||   | |||||| || |||  | |||||||   | ||  |  |  | |||  ||
Db        313 TGCTTCCATTTTTTTCATTAACTTAAGAGAAATAGTATTCAATCCTTTGGTGAAATTCGA 254

Qy        107 GTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
                | |  || |    ||      ||  | |||| |||| ||     ||  | ||||
Db        253 TGTCTACCAAATGTGTGTCTGTCTAGTACTGATTACAAGATTATAAAAATATAAGT 198

Search completed: March  7, 2006, 13:34:30
Job time : 534.953 secs
```

```
                  GenCore version 5.1.7
         Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:          March  7, 2006, 12:18:50 ; Search time 1775.25 Seconds
                                            (without alignments)
                                            237.125 Million cell updates/sec

Title:           US-10-782-899-1_COPY_1_192
Perfect score:   192
Sequence:        1 gctgggcgcggcggccatga..........tgacccaagaaggatgcttt 192

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        7218535 seqs, 1096242582 residues

Total number of hits satisfying chosen parameters:      14437070

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_New:*
                 1:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
                 7:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
                 9:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
                 11:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
                 12:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
                 13:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                    %
Result              Query
  No.    Score      Match Length DB ID                        Description
 ------------------------------------------------------------------------------
c   1     31.6     16.5   1437  6 US-09-925-065A-691107     Sequence 691107,
c   2     31.6     16.5   1437  6 US-09-925-065A-691108     Sequence 691108,
    3     31.2     16.2   1485  6 US-09-925-065A-49312      Sequence 49312, A
    4     31.2     16.2   1485  6 US-09-925-065A-49313      Sequence 49313, A
c   5     31.2     16.2  96109 12  US-11-124-368A-2895       Sequence 2895, Ap
    6       31     16.1   1771  6 US-09-925-065A-65588      Sequence 65588, A
c   7     30.8     16.0    639  6 US-09-925-065A-170277     Sequence 170277,
c   8     30.8     16.0    639  6 US-09-925-065A-170278     Sequence 170278,
c   9     30.6     15.9    572  6 US-09-925-065A-308556     Sequence 308556,
c  10     30.6     15.9    572  6 US-09-925-065A-308557     Sequence 308557,
c  11     30.6     15.9    601  6 US-09-925-065A-396377     Sequence 396377,
   12     30.6     15.9    606  6 US-09-925-065A-599874     Sequence 599874,
   13     30.6     15.9    670  6 US-09-925-065A-905740     Sequence 905740,
c  14     30.6     15.9    670  6 US-09-925-065A-912520     Sequence 912520,
   15     30.4     15.8    528  6 US-09-925-065A-149916     Sequence 149916,
c  16     30.4     15.8    774  8 US-10-750-185-44581       Sequence 44581, A
c  17     30.4     15.8    774  8 US-10-750-623-44581       Sequence 44581, A
   18     30.4     15.8   1569  8 US-10-793-626-603         Sequence 603, App
   19     30.4     15.8   3204  8 US-10-793-626-3485        Sequence 3485, Ap
   20     30.2     15.7    583  6 US-09-925-065A-460994     Sequence 460994,
   21       30     15.6   1296  6 US-09-925-065A-684507     Sequence 684507,
   22       30     15.6   1296  6 US-09-925-065A-684508     Sequence 684508,
   23     29.8     15.5    568  6 US-09-925-065A-13302      Sequence 13302, A
c  24     29.8     15.5    591  6 US-09-925-065A-51223      Sequence 51223, A
   25     29.8     15.5    595  6 US-09-925-065A-463323     Sequence 463323,
   26     29.6     15.4    573  6 US-09-925-065A-355910     Sequence 355910,
c  27     29.6     15.4   1824 12  US-11-024-959-157         Sequence 157, App
c  28     29.6     15.4   1824 12  US-11-024-959-160         Sequence 160, App
c  29     29.4     15.3    340  6 US-09-925-065A-140508     Sequence 140508,
   30     29.4     15.3    550  6 US-09-925-065A-398346     Sequence 398346,
   31     29.4     15.3    550  6 US-09-925-065A-398347     Sequence 398347,
   32     29.4     15.3    627  6 US-09-925-065A-509858     Sequence 509858,
c  33     29.4     15.3  16139 12  US-11-124-367A-5015       Sequence 5015, Ap
   34     29.2     15.2    557  6 US-09-925-065A-795197     Sequence 795197,
   35     29.2     15.2    560  6 US-09-925-065A-797051     Sequence 797051,
   36     29.2     15.2    592  6 US-09-925-065A-616829     Sequence 616829,
   37     29.2     15.2    592  6 US-09-925-065A-616830     Sequence 616830,
   38     29.2     15.2    592  6 US-09-925-065A-616831     Sequence 616831,
   39     29.2     15.2    592  6 US-09-925-065A-616832     Sequence 616832,
   40     29.2     15.2    599  6 US-09-925-065A-795196     Sequence 795196,
   41     29.2     15.2    638  6 US-09-925-065A-613398     Sequence 613398,
   42     29.2     15.2   2937 12  US-11-080-991-45          Sequence 45, Appl
   43     29.2     15.2   3077 12  US-11-186-284-150         Sequence 150, App
   44     29.2     15.2   3213  7 US-10-921-793-83          Sequence 83, Appl
   45     29.2     15.2   3213 12  US-11-186-284-148         Sequence 148, App


                                 ALIGNMENTS


RESULT 1
US-09-925-065A-691107/c
; Sequence 691107, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 691107
;   LENGTH: 1437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
;   LOCATION: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
;   LOCATION: 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,
;   LOCATION: 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
;   LOCATION: 61, 62
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
;   LOCATION: 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
;   LOCATION: 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102
;   OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691107
```

```
  Query Match             16.5%;  Score 31.6;  DB 6;  Length 1437;
  Best Local Similarity   55.5%;  Pred. No. 9.4;
  Matches   61;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         53 GTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTC 112
              |||| | |  ||  |    | |  |||| |||||||| |  ||    |||    || ||
Db        479 GTCTTTGTGAAAGCCAAATGTGATAATACATGTGAAAGCATTTTGTGAACTAAAGTGTTT 420

Qy        113 GCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
              | | |||||  |   |    |||  ||| |   |||||  ||| | | ||
Db        419 GTACAAACAACCTAAATGTCCATCAATGGCTGAATGGATAAAGAAGAGGT 370


RESULT 2
US-09-925-065A-691108/c
; Sequence 691108, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 691108
;   LENGTH: 1437
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
;   LOCATION: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
;   LOCATION: 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,
;   LOCATION: 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
;   LOCATION: 61, 62
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
;   LOCATION: 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
;   LOCATION: 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102
;   OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691108

  Query Match             16.5%;  Score 31.6;  DB 6;  Length 1437;
  Best Local Similarity   55.5%;  Pred. No. 9.4;
  Matches   61;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         53 GTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTC 112
              |||| | |  ||  |    | |  |||| |||||||| |  ||    |||    || ||
Db        479 GTCTTTGTGAAAGCCAAATGTGATAATACATGTGAAAGCATTTTGTGAACTAAAGTGTTT 420

Qy        113 GCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGT 162
              | | |||||  |   |    |||  ||| |   |||||  ||| | | ||
Db        419 GTACAAACAACCTAAATGTCCATCAATGGCTGAATGGATAAAGAAGAGGT 370


RESULT 3
US-09-925-065A-49312
; Sequence 49312, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49312
;   LENGTH: 1485
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-49312

  Query Match             16.2%;  Score 31.2;  DB 6;  Length 1485;
  Best Local Similarity   57.0%;  Pred. No. 12;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
              ||||   |||  || |||  |||| |    ||| | |||||   ||| ||| || || ||
Db       1062 AAGTCAATTGCTAACCATTTTCTTCTTATTTTCTTATATTTCTTATAGCTTAGGTGACAT 1121

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAA 119
              | ||  ||| |   || | |  |  |  |  |   | |||
Db       1122 AAATAGGAAGATGTTTGTTACTCTAGCATAATAAAATAAA 1161


RESULT 4
US-09-925-065A-49313
; Sequence 49313, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49313
;   LENGTH: 1485
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-49313

  Query Match             16.2%;  Score 31.2;  DB 6;  Length 1485;
  Best Local Similarity   57.0%;  Pred. No. 12;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
              ||||   |||  || |||  |||| |    ||| | |||||   ||| ||| || || ||
```

```
Db       1062 AAGTCAATTGCTAACCATTTTCTTCTTATTTTCTTATATTTCTTATAGCTTAGGTGACAT 1121

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAA 119
              | ||  ||| |   || | |  |  |  |  |   | |||
Db       1122 AAATAGGAAGATGTTTGTTACTCTAGCATAATAAAATAAA 1161

RESULT 5
US-11-124-368A-2895/c
; Sequence 2895, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  James J. Devlin
;  APPLICANT:  May Luke
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Vascular Diseases, Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001524
;  CURRENT APPLICATION NUMBER: US/11/124,368A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,845
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/625,936
;  PRIOR FILING DATE: 2004-11-09
;  NUMBER OF SEQ ID NOS: 21112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2895
;   LENGTH: 96109
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-2895

  Query Match             16.2%;  Score 31.2;  DB 12;  Length 96109;
  Best Local Similarity   53.2%;  Pred. No. 43;
  Matches   66;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy         63 AATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACAT 122
              ||   | ||  || | ||   |  |||||||||  || | || || |   | ||  | ||
Db      63419 AAGGACCTCAAGGGAGTAATAGAAAAAACCCTTCCAACAAGCTGGATGATGAAGGGAGAT 63360

Qy        123 GCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCTGACCCAAGA 182
              | |  |||         ||| | | || ||   || || ||  |||  |  | | |||
Db      63359 GAGACAGTGGACAGAGAGACCAGAGGGGCTCCTCATAATTGTGTGTGACTAGCCACAACT 63300

Qy        183 AGGA 186
              | ||
Db      63299 AAGA 63296

RESULT 6
US-09-925-065A-65588
; Sequence 65588, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 65588
```

```
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-65588

  Query Match             16.1%;  Score 31;  DB 6;  Length 1771;
  Best Local Similarity   53.8%;  Pred. No. 15;
  Matches   64;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy         55 CTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGC 114
              |||  |  |||  |||| || ||||| || ||| ||| | ||| |||        |||
Db       1186 CTAACTAAAATGTCTTCAGGCAAATAAATTTGACAACTCATATGAAATAGACAAATTCCT 1245

Qy        115 AGAAACATGCGCTAGTATCATTGATGACAACATGGACTAAGCAAAAGTGCTTGTCCCCT 173
              ||||| | || |  | |  ||  |  | ||        | |  |||    |||   |||
Db       1246 AGAAAAACGCACATGAAAAATAAAATATAATCATATAAATGTTAAAAATATTGAATCCT 1304

RESULT 7
US-09-925-065A-170277/c
; Sequence 170277, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170277
;   LENGTH: 639
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-170277

  Query Match             16.0%;  Score 30.8;  DB 6;  Length 639;
  Best Local Similarity   55.7%;  Pred. No. 13;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         43 TGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAAC 102
              | ||||||| || || |||   | | ||| |  ||||  |  |  |     |  |  ||
Db        616 TTTCTGATTTGTGTAATTTTTCTTCTTTCTGAAAAATTTAGTTTTACCATATCTTGTAAT 557

Qy        103 GCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
              || | |||  | ||| | || | ||  | |  || ||  ||| | |
Db        556 GCAGATTTCTGGAGACAAATTCTCTGTTTTTGTTTATCTCAAAAAG 511

RESULT 8
US-09-925-065A-170278/c
; Sequence 170278, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
```

```
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170278
;   LENGTH: 639
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-170278

  Query Match             16.0%;  Score 30.8;  DB 6;  Length 639;
  Best Local Similarity   55.7%;  Pred. No. 13;
  Matches   59;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         43 TGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAAC 102
              | ||||||| || || |||   | | ||| |  ||||  |  |  |      |  |  ||
Db        616 TTTCTGATTTGTGTAATTTTTCTTCTTTCTGAAAAATTTAGTTTTACCATATCTTGTAAT 557

Qy        103 GCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATG 148
              || | |||  | ||| | || | ||  | |  || ||  ||| | |
Db        556 GCAGATTTCTGGAGACAAATTCTCTGTTTTTGTTTATCTCAAAAAG 511


RESULT 9
US-09-925-065A-308556/c
; Sequence 308556, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 308556
;   LENGTH: 572
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-308556

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 572;
  Best Local Similarity   50.3%;  Pred. No. 14;
  Matches   75;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              |||  | || | | | | |    ||||||| |||||  | |   |     |   |||
Db        397 CATGGTTTTATTTAAATTGAGACTTTTCAAAACCTTGTGTGTTTTTATATTACCAACATA 338

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               ||  ||  |  |  ||  |||| |  |  | |||      | |  | ||    ||  |
Db        337 AATTGAAAACCAGCCTTACCAGACATTTAAGATAGCCATGATTAATATAAACATGAAAAT 278

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAA 183
                |||| || ||| ||  | ||||| | |
Db        277 TTAAAAATGTTTGCCCATTTACCCATGTA 249


RESULT 10
US-09-925-065A-308557/c
; Sequence 308557, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 308557
;   LENGTH: 572
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-308557

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 572;
  Best Local Similarity   50.3%;  Pred. No. 14;
  Matches   75;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy         35 CATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCT 94
              |||  | || | | | | |    ||||||| |||||  | |   |     |   |||
Db        397 CATGGTTTTATTTAAATTGAGACTTTTCAAAACCTTGTGTGTTTTTATATTACCAACATA 338

Qy         95 TATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGATGACAACATGGACTAA 154
               ||  ||  |  |  ||  |||| |  |  | |||      | |  | ||    ||  |
Db        337 AATTGAAAACCAGCCTTACCAGACATTTAAGATAGCCATGATTAATATAAACATGAAAAT 278

Qy        155 GCAAAAGTGCTTGTCCCCTGACCCAAGAA 183
                |||| || ||| ||  | ||||| | |
Db        277 TTAAAAATGTTTGCCCATTTACCCATGTA 249


RESULT 11
US-09-925-065A-396377/c
; Sequence 396377, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 396377
;   LENGTH: 601
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-396377

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 601;
  Best Local Similarity   58.1%;  Pred. No. 15;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
               || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db         544 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 485

Qy         116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
                ||| ||| |||   || ||  ||||  | |||
Db         484 CAAAGATGAGCTGAAATAATACATGAATAAATG 452


RESULT 12
US-09-925-065A-599874
; Sequence 599874, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 599874
;   LENGTH: 606
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-599874

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 606;
  Best Local Similarity   58.1%;  Pred. No. 15;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          56 TATTTTCAATACCTTCGGGGAAATAGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCA 115
               || ||||| || |||  || |  |  | | |||||     ||||| |     | ||  |
Db          63 TACTTTCATTAACTTTAGGAAGCTGAAGGGGAAAAAATCAATAAAGCAGCTCTCTTGACT 122

Qy         116 GAAACATGCGCTAGTATCATTGATGACAACATG 148
                ||| ||| |||   || ||  ||||  | |||
Db         123 CAAAGATGAGCTGAAATAATACATGAATAAATG 155


RESULT 13
US-09-925-065A-905740
; Sequence 905740, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 905740
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-905740

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 670;
  Best Local Similarity   60.0%;  Pred. No. 15;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          21 AGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATA 80
               | |  || ||||   ||| ||||||    | | | | ||||  ||||   |    |||||
Db         541 ACTAACTAGACAGAGATATTCTTGTAATTTACTTTTGTTTTACATACTCACATATAAATA 600

Qy          81 GATGTGAAAACCCTTATAAAACGCG 105
                |||  | |||   |||| |||| |
Db         601 TATGATATAACATGTATAGAACGAG 625


RESULT 14
US-09-925-065A-912520/c
; Sequence 912520, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 912520
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-912520

  Query Match             15.9%;  Score 30.6;  DB 6;  Length 670;
  Best Local Similarity   60.0%;  Pred. No. 15;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          21 AGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAATA 80
               | |  || ||||   ||| ||||||    | | | | ||||  ||||   |    |||||
Db         130 ACTAACTAGACAGAGATATTCTTGTAATTTACTTTTGTTTTACATACTCACATATAAATA 71

Qy          81 GATGTGAAAACCCTTATAAAACGCG 105
                |||  | |||   |||| |||| |
Db          70 TATGATATAACATGTATAGAACGAG 46


RESULT 15
US-09-925-065A-149916
; Sequence 149916, Application US/09925065A
```

```
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149916
;   LENGTH: 528
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-149916

  Query Match             15.8%;  Score 30.4;  DB 6;  Length 528;
  Best Local Similarity   50.0%;  Pred. No. 16;
  Matches   73;  Conservative    1;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         20 AAGTGGCTTGACAAGCATAATCTTGTCTGATTCGTCTATTTTCAATACCTTCGGGGAAAT 79
              ||||  ||| | ||| ||||| |    || |   |  |||| || |  | | |   |
Db        365 AAGTTTCTTAAGAAGAATAATTTCTGTTGTTGTTTTCATTTACATTTACCTAAGTCGATA 424

Qy         80 AGATGTGAAAACCCTTATAAAACGCGGGTTTTCGCAGAAACATGCGCTAGTATCATTGAT 139
                |||| |     |||| |  |   |  |  ||  |  || ||||   || || | ||||
Db        425 GAATGTAATCTTTCTTACATGATTTGTTTAGTCTAATCAAAATGCTGAAGCATAAGTGAT 484

Qy        140 GACAACATGGACTAAGCAAAAGTGCT 165
               |:       | | |   ||| || |
Db        485 AAYCCTCCATAATGATACAAACTGAT 510


Search completed: March  7, 2006, 13:20:42
Job time : 1776.25 secs
```

```
                  GenCore version 5.1.7
         Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:          March  7, 2006, 11:25:13 ; Search time 1196.77 Seconds
                                            (without alignments)
                                            6127.182 Million cell updates/sec

Title:           US-10-782-899-1_COPY_472_600
Perfect score:   129
Sequence:        1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:      11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_in:*
                 3:  gb_env:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pr:*
                 9:  gb_ro:*
                 10:  gb_sts:*
                 11:  gb_sy:*
                 12:  gb_un:*
                 13:  gb_vi:*
                 14:  gb_htg:*
                 15:  gb_pl:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                 %
Result           Query
  No.   Score    Match Length DB  ID                            Description
  -------------------------------------------------------------------------------
     1     129   100.0    600  1  AB072446                      AB072446 Bifidobac
     2     129   100.0    600  6  BD144660                      BD144660 Drug for
c    3     129   100.0  56640  1  AE014295_22                   Continuation (23 o
c    4     129   100.0 156638  6  AX492788                      AX492788 Sequence
c    5     129   100.0 156646  6  AX553955                      AX553955 Sequence
c    6    32.6    25.3 295650  1  BX294151                      BX294151 Pirellula
c    7      32    24.8 194439  2  CNS07EGA                      AL590443 chromosom
     8    31.8    24.7   3048 15   AK119844                      AK119844 Oryza sat
c    9    31.4    24.3  90417 14   AC138542                      AC138542 Magnaport
c   10    31.4    24.3  92770 14   AC138544                      AC138544 Magnaport
    11    31.4    24.3 128428 14   AC144889                      AC144889 Bos tauru
    12    31.4    24.3 207821 14   AC093397                      AC093397 Bos tauru
    13      31    24.0   5832  1  AF105113                      AF105113 Streptoco
    14      31    24.0  18617  1  CR931675                      CR931675 Streptoco
    15      31    24.0  18754  1  AF094575                      AF094575 Streptoco
c   16    30.8    23.9 222330 14  AC096255                       AC096255 Rattus no
c   17    30.8    23.9 246369 14  AC125707                       AC125707 Rattus no
    18    30.8    23.9 302050  1  AL935257                      AL935257 Lactobaci
```

```
    19    30.6    23.7    821 10  BV633898                      BV633898 S216P6123
    20    30.6    23.7  26309  6  CQ363747                      CQ363747 Sequence
    21    30.6    23.7 110000  1  AE017283_21                   Continuation (22 o
    22    30.6    23.7 223086 14  AC130043                       AC130043 Rattus no
    23    30.6    23.7 229616 14  AC156070                       AC156070 Bos tauru
c   24      30    23.3    960  1  AY343890                      AY343890 Streptomy
    25      30    23.3 236921 14  AC094151                       AC094151 Rattus no
c   26    29.8    23.1 100326  9  AL669928                      AL669928 Mouse DNA
    27    29.8    23.1 110000 15  AP008214_161                   Continuation (162
    28    29.8    23.1 151080 15  AP005726                       AP005726 Oryza sat
c   29    29.8    23.1 153065 14  AP005094                       AP005094 Oryza sat
c   30    29.6    22.9    436  3  DQ094234                      DQ094234 Unculture
c   31    29.6    22.9   2740 15  AK110305                       AK110305 Oryza sat
c   32    29.6    22.9  56276  7  AY129336                      AY129336 Mycobacte
    33    29.6    22.9 110547  5  BX276099                      BX276099 Zebrafish
c   34    29.6    22.9 125020  8  AF429315                      AF429315 Homo sapi
c   35    29.6    22.9 164179  8  AC116901                      AC116901 Homo sapi
c   36    29.6    22.9 174118  8  AC024405                      AC024405 Homo sapi
    37    29.6    22.9 185341 14  AC069403                       AC069403 Homo sapi
    38    29.6    22.9 230664 14  AC097977                       AC097977 Rattus no
    39    29.4    22.8    870  6  AR480420                      AR480420 Sequence
    40    29.4    22.8    885  6  AR586187                      AR586187 Sequence
    41    29.4    22.8   3520  6  CQ412044                      CQ412044 Sequence
    42    29.4    22.8  16853  1  AF246897                      AF246897 Streptoco
c   43    29.4    22.8  37083  8  AC092021                      AC092021 Homo sapi
    44    29.4    22.8 115469  8  AL356219                      AL356219 Human DNA
    45    29.4    22.8 188183  8  AC105921                      AC105921 Homo sapi


                                  ALIGNMENTS


RESULT 1
AB072446
LOCUS       AB072446                 600 bp    DNA     linear   BCT 15-JAN-2003
DEFINITION  Bifidobacterium longum biovar Longum hup gene for histone-like HU
            protein HB1, complete cds.
ACCESSION   AB072446
VERSION     AB072446.1  GI:19918953
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Takeuchi,A., Matsumura,H. and Kano,Y.
  TITLE     Cloning and expression in Escherichia coli of a gene, hup, encoding
            the histone-like protein HU of Bifidobacterium longum
  JOURNAL   Biosci. Biotechnol. Biochem. 66 (3), 598-603 (2002)
   PUBMED   12005055
REFERENCE   2  (bases 1 to 600)
  AUTHORS   Kano,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) Yasunobu Kano, Kyoto Pharmaceutical
            University, Institute of Molecular and Cellular Biology for
            Pharmaceutical Sciences; 1, Shichono-cho, Misasagi, Yamashina-ku,
            Kyoto 6078412, Japan (E-mail:ykano@mb.kyoto-phu.ac.jp,
            Tel:81755954719, Fax:81755954799)
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1679"
     RBS             180..185
     gene            193..474
                     /gene="hup"
     CDS             193..474
                     /gene="hup"
                     /codon_start=1
                     /transl_table=11
                     /product="histone-like HU protein HB1"
                     /protein_id="BAB88690.1"
                     /db_xref="GI:19918954"
```

```
                     /translation="MAYNKSDLVSKIAQKSNLTKAQAEAAVNAFQDVFVEAMKSGEGL
                     KLTGLFSAERVKRPARTGRNPRTGEQIDIPASYGVRISAGSLLKKAVTE"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 1;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
BD144660
LOCUS       BD144660                 600 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Drug for gene therapy using anaerobic bacteria.
ACCESSION   BD144660
VERSION     BD144660.1  GI:27850418
KEYWORDS    JP 2002097144-A/1.
SOURCE      Bifidobacterium longum
  ORGANISM  Bifidobacterium longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Fujimori,M., Taniguchi,S., Amano,J., Yazawa,K., Kano,Y.,
            Nakamura,T. and Sasaki,T.
  TITLE     Drug for gene therapy using anaerobic bacteria
  JOURNAL   Patent: JP 2002097144-A 1 02-APR-2002;
            JUN AMANO,MINORU FUJIMORI
COMMENT     OS   Bifidobacterium longum
            PN   JP 2002097144-A/1
            PD   02-APR-2002
            PF   21-SEP-2001 JP 2001290187
            PI   MINORU FUJIMORI,SHUNICHIRO TANIGUCHI,JUN
            AMANO,KAZUTORA YAZAWA,
            PI   YASUMASA KANO,TOSHIYUKI NAKAMURA,TAKAYUKI SASAKI PC
            A61K35/74,A61K31/396,A61K31/7068,A61K31/711,A61K38/00,A61K38/ PC
            43,A61K48/00,
            PC   A61P35/00,C12N1/21,C12N15/09//(C12N1/21,C12R1:01),C12N15/00,
            PC   A61K37/02,
            PC   A61K37/48
            CC   Drug for gene therapy using anaerobic bacteria FH   Key
                 Location/Qualifiers
            FT   CDS                 (193)..(471).
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Bifidobacterium longum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:216816"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 3
AE014295_22/c
WPCOMMENT
  Sequence split into 23 fragments  LOCUS AE014295 Accession AE014295
     Fragment Name      Begin      End
     AE014295_00            1   110000
     AE014295_01       100001   210000
     AE014295_02       200001   310000
     AE014295_03       300001   410000
     AE014295_04       400001   510000
     AE014295_05       500001   610000
     AE014295_06       600001   710000
     AE014295_07       700001   810000
     AE014295_08       800001   910000
     AE014295_09       900001  1010000
     AE014295_10      1000001  1110000
     AE014295_11      1100001  1210000
     AE014295_12      1200001  1310000
     AE014295_13      1300001  1410000
     AE014295_14      1400001  1510000
     AE014295_15      1500001  1610000
     AE014295_16      1600001  1710000
     AE014295_17      1700001  1810000
     AE014295_18      1800001  1910000
     AE014295_19      1900001  2010000
     AE014295_20      2000001  2110000
     AE014295_21      2100001  2210000
     AE014295_22      2200001  2256640
Continuation (23 of 23) of AE014295 from base 2200001 (AE014295 Bifidobacterium longum

  Query Match             100.0%;  Score 129;  DB 1;  Length 56640;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41386 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 41327

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41326 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 41267

Qy        121 AGTTCAGCG 129
              |||||||||
Db      41266 AGTTCAGCG 41258


RESULT 4
AX492788/c
LOCUS       AX492788               156638 bp    DNA     linear   PAT 26-SEP-2002
DEFINITION  Sequence 1106 from Patent EP1227152.
ACCESSION   AX492788
VERSION     AX492788.1  GI:23338471
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   None.
  TITLE     Bacterial strain and genome of bifidobacterium
  JOURNAL   Patent: EP 1227152-A 1106 31-JUL-2002;
            Societe des Produits Nestle S.A. (CH)
FEATURES             Location/Qualifiers
     source          1..156638
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="1106 seq in place of 1099 because seq 1 (2.256.638
```

```
                     bases splitted
                     in 7 more sequences.
                     seq 0001: from 0.000.001 to 0.349.980 length: 349980
                     seq 1100: from 0.300.001 to 0.649.980 length: 349980
                     seq 1101: from 0.600.001 to 0.949.980 length: 349980
                     seq 1102: from 0.900.001 to 1.249.980 length: 349980
                     seq 1103: from 1.200.001 to 1.549.980 length: 349980
                     seq 1104: from 1.500.001 to 1.849.980 length: 349980
                     seq 1105: from 1.800.001 to 2.149.980 length: 349980
                     seq 1106: from 2.100.001 to 2.256.638 length: 156638"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 6;  Length 156638;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141691 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141632

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141631 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141572

Qy        121 AGTTCAGCG 129
              |||||||||
Db     141571 AGTTCAGCG 141563


RESULT 5
AX553955/c
LOCUS       AX553955               156646 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 9 from Patent WO02074798.
ACCESSION   AX553955
VERSION     AX553955.1  GI:25897908
KEYWORDS    .
SOURCE      Bifidobacterium longum biovar Longum
  ORGANISM  Bifidobacterium longum biovar Longum
            Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
            Bifidobacteriaceae; Bifidobacterium.
REFERENCE   1
  AUTHORS   Arigoni,F., Delley,M., Mollet,B., Pridmore,R.D., Schell,M.A.,
            Pohl,T.G. and Zwahlen,M.C.
  TITLE     The genome of a bifidobacterium
  JOURNAL   Patent: WO 02074798-A 9 26-SEP-2002;
            SOCIETE DES PRODUITS NESTLE S.A. (CH)
FEATURES             Location/Qualifiers
     source          1..156646
                     /organism="Bifidobacterium longum biovar Longum"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1679"
                     /note="seq 1: original length 2256646 splitted into
                     seq 1: from 0.000.001 to 0.349.980
                     seq 3: from 0.300.001 to 0.649.980
                     seq 4: from 0.600.001 to 0.949.980
                     seq 5: from 0.900.001 to 1.249.980
                     seq 6: from 1.200.001 to 1.549.980
                     seq 7: from 1.500.001 to 1.849.980
                     seq 8: from 1.800.001 to 2.149.980
                     seq 9: from 2.100.001 to 2.256.646"
ORIGIN

  Query Match             100.0%;  Score 129;  DB 6;  Length 156646;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-30;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141699 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141640

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141639 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141580
```

```
Qy        121 AGTTCAGCG 129
              |||||||||
Db     141579 AGTTCAGCG 141571


RESULT 6
BX294151/c
LOCUS       BX294151               295650 bp    DNA     linear   BCT 17-APR-2005
DEFINITION  Pirellula sp. strain 1 complete genome; segment 19/24.
ACCESSION   BX294151 BX119912
VERSION     BX294151.1  GI:32446812
KEYWORDS    complete genome.
SOURCE      Rhodopirellula baltica SH 1
  ORGANISM  Rhodopirellula baltica SH 1
            Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
            Planctomycetaceae; Rhodopirellula.
REFERENCE   1  (bases 1 to 295650)
  AUTHORS   Glockner,F.O., Kube,M., Bauer,M., Teeling,H., Lombardot,T.,
            Ludwig,W., Gade,D., Beck,A., Borzym,K., Heitmann,K., Rabus,R.,
            Schlesner,H., Amann,R. and Reinhardt,R.
  TITLE     Complete genome sequence of the marine planctomycete Pirellula sp.
            strain 1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
   PUBMED   12835416
REFERENCE   2  (bases 1 to 295650)
  AUTHORS   Kube,M., Borzym,K., Heitmann,K., Klages,S., Marquardt,I.,
            Lehrack,S., Beck,A., Pawlik,R., Reinhardt,R., Gloeckner,F.O.,
            Bauer,M., Teeling,H., Lombardot,T., Ludwig,W., Gade,D., Rabus,R.,
            Schlesner,H. and Amann,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2003) Max Planck Institute for Molecular
            Genetics, proScience Ihnestrasse 73, D-14195 Berlin, Germany Max
            Planck Institute for Marine Microbiology Celsiusstrasse 1, D-28359
            Bremen, Germany
COMMENT     This project was carried out by
            *Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
            Planck Institute for Marine Microbiology, Bremen, Germany; in the
            fremework of the REGX-project, http://www.regx.de -------------
            Genome Center
                 Center: Max Planck Institute for Molecular Genetics
                 Center code: MPIMG
            -------------- Summary Statistics
                 Sequencing vector: pUC19; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 7142841 bases at least Q40
                 Consensus quality: 7145138 bases at least Q30
                 Consensus quality: 7145484 bases at least Q20
                 Quality coverage: 8.03
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or coveredby high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; allregions were covered by at least
            one plasmid Sequence; assembly was additionally confirmed by long
            range pcr and cosmid end sequences.
            --------------
            See http://www.micro-genomes.mpg.de/pirellula/ for more information
            including minimal tiling path from a set of  220 cosmids out of
            908. See the misc_feature tag below for the boundaries of the MTP
            cosmids. -------------- Annotation
                 Center: Max Planck Institute for Marine Microbiology
                         Celsiusstrasse 1, D-28359 Bremen, Germany.
                 Center Code: MPIMM
                 Email: fog@mpi-bremen.de
                 Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580
            --------------
            Three different programs (Glimmer, Critica and Orpheus) were used
            for ORF-prediction. A nonredundant list of ORFs was generated by
            suitable parsing of the results.
            Automated annotation was done with the software package Pedant Pro
```

```
            (http://www.biomax.de). All ORF predictions and annotations were
            manually corrected by considering all results of the different
            tools applied. See http://www.regx.de for more information and
            access to supplementary information.
            --------------.
FEATURES             Location/Qualifiers
     source          1. .295650
                     /organism="Rhodopirellula baltica SH 1"
                     /mol_type="genomic DNA"
                     /strain="1"
                     /db_xref="taxon:243090"
     gene            complement(414. .1508)
                     /gene="pknH"
                     /locus_tag="RB10133"
     CDS             complement(414. .1508)
                     /gene="pknH"
                     /locus_tag="RB10133"
                     /EC_number="2.7.1.-"
                     /function="unspecified signal transduction; unspecified
                     kinase or ATP dependent regulatory protein; cellular
                     communication/signal transduction"
                     /note="best DB hits: BLAST: swissprot:Q11053; PKNH_MYCTU
                     PROBABLE SERINETHREONINE-PROTEIN; E=3e-35 embl:CAB94054.1;
                     (AL358672) putative serinethreonine-protein; E=1e-34
                     embl:CAA10713.1; (AJ132604) hypothetical protein
                     [Lactococcus; E=2e-33 COG: Rv1266c; COG0515
                     Serine/threonine protein kinases; E=3e-36 PFAM: PF00069;
                     Protein kinase domain; E=7.6e-54"
                     /codon_start=1
                     /transl_table=11
                     /product="probable serine/threonine-protein kinase pknH"
                     /protein_id="CAD78719.1"
                     /db_xref="GI:32446813"
                     /db_xref="GOA:Q7UFG3"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR008271"
                     /db_xref="UniProt/TrEMBL:Q7UFG3"
                     /translation="MMPHLRLTDFELGAVLGVGTVGTVYDGKILDDVEVHPAAEAIRG
                     QDLAVKKLHPAVSQDDLIQARFRREMVILERLQHPNIIGYFGGGSEDGQLFYVMERVD
                     GGTIKDLLETNGALAWPVVVDVARQVCSALQCAHNHGVIHRDLKPGNLFLTRDAHVKL
                     GDFGIARDQHSSDLTSQGLTVGTHAYMAPEQITGDEAISGKADLYALGCVLFEMLANR
                     KVFAGENFAQLFEQHLRTKAPTIASIVPDVPPELSQVIADCLEKSPDDRPFNARSVQG
                     VMIEIGEKYNLSAPPDSPVGPNHSETGNAPKDVSADSVTEKGRRLLEEQIHYRLGGTS
                     REPVSAGKVSLIVVAIIVLIVLAVSLSGGS"
     gene            complement(1529. .1843)
                     /locus_tag="RB10134"
     CDS             complement(1529. .1843)
                     /locus_tag="RB10134"
                     /function="ribosomal proteins"
                     /note="best DB hits: BLAST: gb:AAK05177.1; AE006340_10
                     (AE006340) 50S ribosomal protein L21; E=4e-10 pir:F82920;
                     ribosomal protein L21 UU212 [imported] - Ureaplasma;
                     E=3e-11 gb:AAC35614.1; (AF041468) ribosomal protein L21
                     [Guillardia; E=3e-10 COG: Rv2442c; COG0261 Ribosomal
                     protein L21; E=3e-07 PFAM: PF00829; Ribosomal prokaryotic
                     L21 prot; E=2.2e-36"
                     /codon_start=1
                     /transl_table=11
                     /product="50S ribosomal protein L21"
                     /protein_id="CAD78720.1"
                     /db_xref="GI:32446814"
                     /db_xref="GOA:Q7UFG2"
                     /db_xref="InterPro:IPR001787"
                     /db_xref="UniProt/TrEMBL:Q7UFG2"
                     /translation="MYAIFVDGGRQYRVEPGMELDVDYRDIAAGENLKFETVLAVGAE
                     DGLKLGAPTLDGVSVSASVLGMSQDKKIYIQKFRRRKHSKKRTGHRQKNTRIRIEEIA
                     GV"
     gene            complement(1767. .2018)
                     /locus_tag="RB10135"
     CDS             complement(1767. .2018)
                     /locus_tag="RB10135"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="hypothetical protein"
                     /protein_id="CAD78721.1"
                     /db_xref="GI:32446815"
                     /db_xref="UniProt/TrEMBL:Q7UFG1"
                     /translation="MLDRPIFGSTDLFTSDDLLDSLALKNQPWSKRRSVVILRDSKIR
                     PSGQFPLHPEPFSRHVCHLCRRWTPIPSRARNGTRRRLP"
     gene            2019. .2177
                     /locus_tag="RB10136"
     CDS             2019. .2177
                     /locus_tag="RB10136"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD78722.1"
                     /db_xref="GI:32446816"
                     /db_xref="UniProt/TrEMBL:Q7UFG0"
                     /translation="MLVLRNRSELSVGYSGAARAARCVTGGLAPNGTYFTPPQGGCEF
                     LRASPVTP"
     gene            complement(2151. .2372)
                     /locus_tag="RB10137"
     CDS             complement(2151. .2372)
                     /locus_tag="RB10137"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein-putative conserved
                     hypothetical protein"
                     /protein_id="CAD78723.1"
                     /db_xref="GI:32446817"
                     /db_xref="InterPro:IPR011481"
                     /db_xref="UniProt/TrEMBL:Q7UFF9"
                     /translation="MSRSRFSGVLYDAPLPKRSLGGGRATPFRRTRGRGPSMENGADC
                     PPPEISLNARFPTPPAARAGFSRRYWRCT"
     gene            complement(2478. .3281)
                     /locus_tag="RB10138"
     CDS             complement(2478. .3281)
                     /locus_tag="RB10138"
                     /function="inner membrane organization"
                     /note="best DB hits: BLAST: pir:T34929; hypothetical
                     protein SC3F9.07 SC3F9.07 - Streptomyces; E=0.018"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein-putative membrane
                     lipoprotein"
                     /protein_id="CAD78724.1"
                     /db_xref="GI:32446818"
                     /db_xref="UniProt/TrEMBL:Q7UFF8"
                     /translation="MRLVTDFFNAFPFLQRNSTSSALMSLVFAGCCLLNVGFESIAEA
                     DEPAEQTSNVSVKIFDTNAPGWRQLDGDDFVRVNGDDQTLTWEGTEALGSGQPIGVTR
                     TNFEVKNFELVIQWKHLKPAGNSGVFAWVPMSALKDLPPNRLPNTGIEVQMLDLDYGR
                     MYTEKTGKPPTWFTSHGDIFAVGKSSMQPFPPLSPDGHRSFPSAETTNPHGEWNQYYV
                     RGINREIRLWVNGVEVSGGRSCSPDEGFLCLESEGSPIRFREIWLRELP"
     gene            complement(3382. .3621)
                     /locus_tag="RB10139"
     CDS             complement(3382. .3621)
                     /locus_tag="RB10139"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD78725.1"
                     /db_xref="GI:32446819"
                     /db_xref="UniProt/TrEMBL:Q7UFF7"
                     /translation="MRHKHDRMKIRGISKNPIDLPSRYRGRVLKPKKRRRSFRSFLRE

  Query Match             25.3%;  Score 32.6;  DB 1;  Length 295650;
  Best Local Similarity   53.5%;  Pred. No. 24;
  Matches   68;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              |||| |    |||| | ||| || ||||| |      |||||| ||||||     | |||
Db     148951 TGACGTGAGCCTCGCATCGAATAGTTCGAACGACGACGACGACGAAGACCAACGTCAAGA 148892

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              || |   | |  ||| |  | | | | |   |     |    | | |  | | ||| |||
```

```
Db      148891 TGTTGTCGTTGATTTGGCGGAGTTTCGGATGCAGTCCGAAAGAGCATTCTCTACCGAACT 148832

Qy         121 AGTTCAG 127
                 ||| |
Db      148831 GTTTCGG 148825


RESULT 7
CNS07EGA/c
LOCUS       CNS07EGA               194439 bp    DNA     linear   INV 16-APR-2005
DEFINITION  chromosome III of strain GB-M1 of Encephalitozoon cuniculi
            (Microspora).
ACCESSION   AL590443
VERSION     AL590443.1  GI:13560062
KEYWORDS    AIDS opportunistic pathogen; Genome sequence; intracellular
            parasite.
SOURCE      Encephalitozoon cuniculi GB-M1
  ORGANISM  Encephalitozoon cuniculi GB-M1
            Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
REFERENCE   1  (bases 1 to 194439)
  AUTHORS   Katinka,M.D., Duprat,S., Cornillot,E., Metenier,G., Thomarat,F.,
            Prensier,G., Barbe,V., Peyretaillade,E., Brottier,P., Wincker,P.,
            Delbac,F., El Alaoui,H., Peyret,P., Saurin,W., Gouy,M.,
            Weissenbach,J. and Vivares,C.P.
  TITLE     Genome sequence and gene compaction of the eukaryote parasite
            Encephalitozoon cuniculi
  JOURNAL   Nature 414 (6862), 450-453 (2001)
   PUBMED   11719806
REFERENCE   2  (bases 1 to 194439)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     Corresponding author : Equipe de Parasitologie Moleculaire et
            Cellulaire, UMR CNRS 6023, Universite Blaise Pascal, 63177 Aubiere
            cedex, France E-mail : Christian.VIVARES@lbp.univ-bpclermont.fr
            Tel : (33) (0)4.73.40.74.57
            Fax : (33) (0)4.73.40.74.55
            Web : http://www.protistes.univ-bpclermont.fr/fr/parasito.htm The
            submitted sequence encompasses the coding core of the chromosome.
            The telomeric and subtelomeric sequences are not included in this
            submission.
FEATURES             Location/Qualifiers
     source          1. .194439
                     /organism="Encephalitozoon cuniculi GB-M1"
                     /mol_type="genomic DNA"
                     /strain="GB-M1"
                     /db_xref="taxon:284813"
                     /chromosome="III"
     gene            550. .1380
                     /gene="ECU03_0010"
     CDS             550. .1380
                     /gene="ECU03_0010"
                     /function="unknown"
                     /note="03_0010, unknown, gene found by Glimmer [Bad Olap 2
                     167 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26148.1"
                     /db_xref="GI:19168629"
                     /db_xref="UniProt/Swiss-Prot:Q8SW85"
                     /translation="MAAPTQLPETAKPKHSNQSEAGPAPLASPTAPMPRPASHLAPMP
                     SDHPDFRSKSARLRCQPPRTNNCGTFKQPPSVAATSRPKPGNPFLQPPTKGTPPPKKK
                     KKNHTEGCHTHEANPEPNTKHTETEPPIISHCPPPHPGPTATPNLLPCPNPTSSFCQN
                     TRDSPSLPPQRPNMVHIPETPLQRRHSPSQIPIRIPPSPYNLSTAKTYYYLQPPYLPI
                     LSHEIHFIVPLLCPYEHHTHNRAARREPATLENNPRHRSIRLNYIPSHHVPHLRRRQF
                     "
     gene            1211. .2020
                     /gene="ECU03_0020"
     CDS             1211. .2020
                     /gene="ECU03_0020"
                     /function="unknown"
                     /note="03_0020, unknown, gene found by Glimmer [Shorter 1
                     167 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26149.1"
                     /db_xref="GI:19168630"
                     /db_xref="UniProt/Swiss-Prot:Q8SW84"
                     /translation="MKSISSYLSFAHMNITHTTEQHAENQPHWKTILDIAPFVSITFP
                     AIMCLIFDEDSFEESPFLRFITLLLPFSYSAVQYALLYTNWKSHNKPEPILHTTLYYT
                     LSLLLLAFTIISILSIIPFSLNEWDHAASFFYPIVLPSFTVPPAYLLSSSYFLVPRQI
                     RLTDTVISILISVCSIVNVLLVFKEFNYYPYSAIISSISVLLQLLSEKHCLFKQSPPS
                     TASSRAAVLILTLILAVLVYTFLGYGAIYILDDHFHLLGKMKSILPSEPHQ"
     gene            complement(3222. .3578)
                     /gene="ECU03_0030"
     CDS             complement(3222. .3578)
                     /gene="ECU03_0030"
                     /function="unknown"
                     /note="03_0030, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26150.1"
                     /db_xref="GI:19168631"
                     /db_xref="UniProt/Swiss-Prot:Q8STE8"
                     /translation="MCFPVGGLLPYLALRSVAYICVFGDSSRAPDLIREAFESPHFKR
                     FDGAGTYKEAKGRCGMRFADVVNGAFGQISDMADKVGKGEPEVWCIWKKRGEVEMLLK
                     VKEYRKGYGSGKRRRR"
     gene            4169. .4975
                     /gene="ECU03_0040"
     CDS             4169. .4975
                     /gene="ECU03_0040"
                     /function="unknown"
                     /note="03_0040, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26151.1"
                     /db_xref="GI:19168632"
                     /db_xref="InterPro:IPR011004"
                     /db_xref="UniProt/Swiss-Prot:Q8SW83"
                     /translation="MNTTHVPEPHRTEQHTENLQHWRKILGIAPIVSIAFPAIMYFIF
                     TKDSFEDSLFLRFITILLSFSYSAVQYAVLLHTNWKSHNKPEGILHTTLYYTLNLLLL
                     AFSIISILSITTLPINKWKNDGGPILFSIFLPPLFMSPAYLLSTSCRLVPGQIGFTDT
                     GINVLIDILTLLCSVGSLLLILEESEYCYCFAIISSILILIRLLGEKLSPEKQSPPPT
                     APWRIAVFVLILIFAALIYAFMMWVSIDILSDHFDLLTKARSTSVSKPGQ"
     gene            5400. .6665
                     /gene="ECU03_0050"
     CDS             5400. .6665
                     /gene="ECU03_0050"
                     /function="unknown"
                     /note="03_0050, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26152.1"
                     /db_xref="GI:19168633"
                     /db_xref="InterPro:IPR001664"
                     /db_xref="UniProt/TrEMBL:Q8SW82"
                     /translation="MEQISRIKTTLDEVDLDYDERAFRGLSVEKLWVRAFWLISDIVK
                     CEKMHPAIACYRNLLDSILEDLEEKGHPEGITAVNAAKDALRALEDVVNRKCQILSLS
                     KKAHGTEFPYYFGAIDGYLKDYPGLLKAFKDALRKVLPHVRGDLKVKVNEMKLEAAMS
                     DFLKVLKKQLDGEKVQERPSTLDIYKYLMNEVNEVEVPLKSVIRDLLGIEGNYNFEAN
                     WSFYSYPRDINEAALYIFQKYFQHDIDEASLSRTIKNELECLEKVISKVPKSWRARLQ
                     RILRDVRTAFDSLSQKLKALRTSCTRVIDDVESNGKSFSDAFSELNGMLREAEKGLSD
                     LKERISDLERVLPSSPPMKLCELLCFMFIAILAIHSFLHLAEDKNSRVKTRWNTALYS
                     LAVAGTIMHQMWVLWGGGGTGMRMGQVMC"
     gene            6782. .7153
                     /gene="ECU03_0060"
     CDS             6782. .7153
                     /gene="ECU03_0060"
                     /function="unknown"
                     /note="03_0060, unknown, gene found by Glimmer [Delay by
                     228 5 109 38]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26153.1"
```

```
                     /db_xref="GI:19168634"
                     /db_xref="UniProt/TrEMBL:Q8SW81"
                     /translation="MCGQRLIQCGMSWKQMCVFCAGHVLLAVACFAEPSKHPLLDTYG
                     RCGMVCIGLTVLIALLLLVSYMDEADGEKASASGSGAVVMNCLFVMSVAVTTFASWVS
                     SRDYFMTCRRGTDDMTGTAAS"
     gene            complement(7642. .7953)
                     /gene="ECU03_0070"
     CDS             complement(7642. .7953)
                     /gene="ECU03_0070"
                     /function="unknown"
                     /note="03_0070, unknown, gene found by Glimmer [Shorter 8
                     242 99]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26154.1"
                     /db_xref="GI:19168635"
                     /db_xref="UniProt/TrEMBL:Q8SW80"
                     /translation="MIERERETVLYFQLWMANAMTERMEMTANVMKRALEKYRSILCM
                     APSGLVLSFQLARNRRTIGNAIDMNTARSLGIVPPPIIELVLGVLLCCVCGVHAGKGE
                     V"
     gene            7712. .8389
                     /gene="ECU03_0080"
     CDS             7712. .8389
                     /gene="ECU03_0080"
                     /function="unknown"
                     /note="03_0080, unknown, gene found by Glimmer [Bad Olap 7
                     242 0]"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26155.1"
                     /db_xref="GI:19168636"
                     /db_xref="UniProt/Swiss-Prot:Q8SW79"
                     /translation="MGGGTIPRDLAVFMSIAFPMVLLFRASWKESTKPEGAIHKILRY
                     FSSALFITFAVISILSVIAFAIHNWKYNTVSLSLSIIFPSLLFPPTYLLSTSCCLVPG
                     QIGFTDTGINILIDILILVHPMANLVSVRRRPYRYPHSAIISSIFLIIIFTPDNLNNG
                     ATAFCSQLSPPYLDLQPIFSVLSASRQHHIPAFLLALACPIMGLIPMIKDSSTIPISR
                     LPLLPSF"
     gene            complement(8576. .10444)
                     /gene="ECU03_0090"
     CDS             complement(8576. .10444)
                     /gene="ECU03_0090"
                     /function="unknown"
                     /note="03_0090, unknown, gene found by Glimmer"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD26156.1"
                     /db_xref="GI:19168637"
                     /db_xref="UniProt/Swiss-Prot:Q8SW78"
                     /translation="MGECPFQTMRMWLVSAIALVDILCTSKIEELVENMGKKIEEALP
                     HEFDSAEKEALKKHLKETAGLGTRLMVPCIFHEDRVVASPTTRYQDIDEEEKGYVEKV
                     IALLPRLLWRSVAYIYVFGNDSWVVNLMEEVFETAPFKKSDAVALYKRARGRLGIRLI
                     DLVNRTFRHNIGMLNRFGQRLAQEAEAKIQEISSSLSDEEKRKEEKMLQIIKEYGESL
                     CTKEKQEEIIRAQEIMCDACACIWERDSNRESFVMETYSRHLYLRMIGSSMDVEEPLL
                     SYIDHRGLIDAYEKYKSIDIVAELIKQVFTERGCISDESVNDAVCGVREREEAEKMRG
                     EEERRKKEEESLRNTLELLRMEEKEKSKSRGKKKKGGKKGSGEVTAKMEEEKKDSEEV
                     EESAEAEVSLEEMAVGGARSKERSSKKKSRSKGHRYKVHKRVLRWTKSAERIKAELDE
                     GSEEKWRNKSIEEIEEQKKVHDIIEVCVLLRSLDANRFFVSTNRYMKDGTERWKMVGV
                     GIFEEGGEKKVGKVEVGLYRDKGEGSVIYHLMFKAMDTEKAGKGARSSFGKGDDVEGL
                     EEGAGELSDMSGFEYPKGVRSEIVKGGDAFKIVYRNPKDTSEVLRSLTVLQKAEVL"
     gene            11401. .12966
                     /gene="ECU03_0100"
     CDS             11401. .12966
                     /gene="ECU03_0100"

  Query Match            24.8%;  Score 32;  DB 2;  Length 194439;
  Best Local Similarity  58.3%;  Pred. No. 37;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||| |||  || |  ||   |    ||||   |||| |||| |  ||||  |  ||
Db     127681 TGACCTTGTGCGTGTCGACATACATGTTTGCATCGTTGACAACAATGTGCCCGTTCTTGA 127622

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              |  || || ||| ||  | ||  || |  |   |||
Db     127621 TCATCAGGATCTCTTCTTGGTTCTGGTCGGCAATGT 127586


RESULT 8
AK119844
LOCUS       AK119844                3048 bp    mRNA    linear   PLN 29-OCT-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-178-D02, full
            insert sequence.
ACCESSION   AK119844
VERSION     AK119844.1  GI:37989467
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
   PUBMED   12869764
REFERENCE   2
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakahama,Y.,
            Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K.,
            Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M.,
            Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C.,
            Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y.,
            Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M.,
            Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K.,
            Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J.,
            Yokomizo,S. and Yoshimura,A.
  TITLE     Collection, mapping, and annotation of 28K full-length cDNA clones
            from japonica rice
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3048)
  AUTHORS   Kikuchi,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 32K full-length cDNA clones from japonica
```

```
              rice.
              URL : http://cdna01.dna.affrc.go.jp/cDNA/
              NIAS Rice Full-Length cDNA Project Team:  Kikuchi,S., Satoh,K.,
              Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
              Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,  Kojima,K., Namiki,T.,
              Ohneda,E., Yahagi,W., Suzuki,K.,  Li,C., Ohtsuki,K., Shishiki,T.,
              Yamamoto,M. and Nakahama,Y.
              FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
              Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M.,
              Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M.,
              Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J.,  Oka,M.,
              Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y.,  Tsunoda,Y., Ueda,M.,
              Xie,Q., Yokomizo,S., Yoshimura,A.,  Matsubara,K. and Murakami,K.
              Genome Exploration Research Group in Riken Genomic Sciences  Center
              and Genome Science Laboratory in Riken: Adachi,J.,  Aizawa,K.,
              Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  Hanagaki,T.,
              Hara,A., Hashizume,W., Hayashida,K.,  Hayatsu,N., Hiramoto,K.,
              Hiraoka,T., Hori,F., Iida,J.,  Imamura,K., Imotani,K., Ishii,Y.,
              Itoh,M., Kagawa,I.,  Kanagawa,S., Katoh,H., Kawai,J.,
              Kishikawa-Hirozane,T.,  Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
              Koya,S.,  Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
              Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M.,  Osato,N.,
              Ota,Y., Saitoh,H., Sakai,C., Sakai,K.,
              Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K.,
              Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M.,
              Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
              Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
              Yasunishi,A. and  Hayashizaki,Y.
FEATURES             Location/Qualifiers
     source          1..3048
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="002-178-D02"
ORIGIN

  Query Match             24.7%;  Score 31.8;  DB 15;  Length 3048;
  Best Local Similarity   61.4%;  Pred. No. 44;
  Matches   51;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              ||  || ||  |  |  |  |||   ||   |||| | | |||| |||       || ||
Db        506 AGTTATCACAGCATGGCTCTCTGCACACCCTGACCGCCAACGAGCTGGCTTCCAGCTGTT 565

Qy         76 TGTTGTGGTGCTGTGACGTGTTG 98
              |||||||||||||||  || |||
Db        566 TGTTGTGGTGCTGTGGTGTTTTG 588


RESULT 9
AC138542/c
LOCUS       AC138542               90417 bp    DNA     linear   HTG 20-MAR-2004
DEFINITION  Magnaporthe grisea chromosome 7 clone 12G19, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC138542
VERSION     AC138542.2  GI:45597487
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 90417)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     The sequence of Magnaporthe grisea chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90417)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) Plant Pathology - Fungal Genomics
            Laboratory, North carolina State University, 840 Main Campus Drive,
            Raleigh, NC 27606, USA
```

```
REFERENCE   3  (bases 1 to 90417)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2004) Center for Integrated Fungal Research,
            North Carolina State University, Plant Pathology - 840 Main Campus
            Dr, Raleigh, NC 27606, USA
COMMENT     On Mar 20, 2004 this sequence version replaced gi:27552536.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1     90417: contig of 90417 bp in length.
FEATURES             Location/Qualifiers
     source          1..90417
                     /organism="Magnaporthe grisea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:148305"
                     /chromosome="7"
                     /clone="12G19"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 14;  Length 90417;
  Best Local Similarity   54.9%;  Pred. No. 58;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
              |||| | || |   |  | | |||| |  ||||    || | | || |||  |||  | |
Db      88659 CTTCCGGTCCTCAGGCCTCCATCGAACCATACTCCATACCAGGTCCTCGAAGGAGTCGTT 88600

Qy         65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
              |  || | ||   | ||| | |||    |||   ||||  ||    | || | |
Db      88599 CTTGGGCGTTCGTTCGTGCTTCTGGTTGGTGCAATCCATGCGCGGAAGTCCTG 88547


RESULT 10
AC138544/c
LOCUS       AC138544               92770 bp    DNA     linear   HTG 20-MAR-2004
DEFINITION  Magnaporthe grisea chromosome 7 clone 15G13, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC138544
VERSION     AC138544.2  GI:45597492
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 92770)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     The sequence of Magnaporthe grisea chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92770)
  AUTHORS   Dean,R.A. Dr, Mitchell,T. Dr, Thon,M. Dr and Brown,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) Plant Pathology - Fungal Genomics
            Laboratory, North carolina State University, 840 Main Campus Drive,
            Raleigh, NC 27606, USA
REFERENCE   3  (bases 1 to 92770)
  AUTHORS   Thon,M.R., Mitchell,T., Brown,D.E., Diener,S., Taro,A., Pan,H. and
            Dean,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2004) Center for Integrated Fungal Research,
            North Carolina State University, Plant Pathology - 840 Main Campus
            Dr, Raleigh, NC 27606, USA
COMMENT     On Mar 20, 2004 this sequence version replaced gi:27552538.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
```

```
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    92770: contig of 92770 bp in length.
FEATURES             Location/Qualifiers
     source          1. .92770
                     /organism="Magnaporthe grisea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:148305"
                     /chromosome="7"
                     /clone="15G13"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 14;  Length 92770;
  Best Local Similarity   54.9%;  Pred. No. 58;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
              |||| | || |   |  | | |||| |  ||||    || | | || |||  |||  | |
Db        468 CTTCCGGTCCTCAGGCCTCCATCGAACCATACTCCATACCAGGTCCTCGAAGGAGTCGTT 409

Qy         65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
              |  || | ||   | ||| | |||    |||   ||||  ||    | ||| |
Db        408 CTTGGGCGTTCGTTCGTGCTTCTGGTTGGTGCAATCCATGCGCGGAAGTCCTG 356


RESULT 11
AC144889
LOCUS       AC144889               128428 bp    DNA     linear   HTG 11-JUN-2003
DEFINITION  Bos taurus clone RP42-541E10, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC144889
VERSION     AC144889.2  GI:31581601
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 128428)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128428)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 128428)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 11, 2003 this sequence version replaced gi:31044309.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cey
                 Center clone name: 541E10

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 127382 bases at least Q40
                 Consensus quality: 127698 bases at least Q30
                 Consensus quality: 127848 bases at least Q20
                 Insert size: 130000; agarose-fp
                 Insert size: 127928; sum-of-contigs
                 Quality coverage: 10.65x in Q20 bases; agarose-fp
                 Quality coverage: 10.82x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    33969: contig of 33969 bp in length
            *   33970    34069: gap of unknown length
            *   34070    53360: contig of 19291 bp in length
            *   53361    53460: gap of unknown length
            *   53461    63539: contig of 10079 bp in length
            *   63540    63639: gap of unknown length
            *   63640   112600: contig of 48961 bp in length
            *  112601   112700: gap of unknown length
            *  112701   118744: contig of 6044 bp in length
            *  118745   118844: gap of unknown length
            *  118845   128428: contig of 9584 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128428
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="RP42-541E10"
                     /clone_lib="RP42"
     misc_feature    1. .48703
                     /note="clone overlaps with GenBank Accession Number
                     AC095020 clone RP42-243E1 (center project name cje)"
     misc_feature    1. .33969
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     gap             33970. .34069
                     /estimated_length=unknown
     misc_feature    34070. .53360
                     /note="assembly_fragment"
     gap             53361. .53460
                     /estimated_length=unknown
     misc_feature    53461. .63539
                     /note="assembly_fragment"
     gap             63540. .63639
                     /estimated_length=unknown
     misc_feature    63640. .112600
                     /note="assembly_fragment"
```

```
     misc_feature    72761. .128428
                     /note="clone overlaps with GenBank Accession Number
                     AC093397 clone RP42-405D15 (center project name cja)"
     gap             112601. .112700
                     /estimated_length=unknown
     misc_feature    112701. .118744
                     /note="assembly_fragment"
     gap             118745. .118844
                     /estimated_length=unknown
     misc_feature    118845. .128428
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 14;  Length 128428;
  Best Local Similarity   53.7%;  Pred. No. 58;
  Matches   65;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          8 CTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGG 67
              || | |||     |  | |||  | | |    |||| |   || | |||  | |   |
Db     111483 CTCCCCGTTTGCCTCCCCTCGGCCCTCAGCTCCGACCAGAGCCTCCACCTGGCTCTCCCA 111542

Qy         68 GGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTTCAG 127
              || ||| |||||||   | ||  | ||  ||||||||  | | |  |||      | |||
Db     111543 GGCCTTGTTGTTGTCTAGTTGCTAAGTCATGTCCAACTCTGTGACCCCGTGGACTTGCAG 111602

Qy        128 C 128
              |
Db     111603 C 111603


RESULT 12
AC093397
LOCUS       AC093397              207821 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Bos taurus clone RP42-405D15, WORKING DRAFT SEQUENCE, 10 ordered
            pieces.
ACCESSION   AC093397
VERSION     AC093397.2  GI:21465397
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 207821)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207821)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 207821)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 19, 2002 this sequence version replaced gi:15277346.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cja
                 Center clone name: 405D15

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 205824 bases at least Q40
                 Consensus quality: 206436 bases at least Q30
                 Consensus quality: 206755 bases at least Q20
                 Insert size: 161000; agarose-fp
                 Insert size: 206921; sum-of-contigs
                 Quality coverage: 11.74x in Q20 bases; agarose-fp
                 Quality coverage: 9.14x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1     22661: contig of 22661 bp in length
            *      22662     22761: gap of unknown length
            *      22762     45605: contig of 22844 bp in length
            *      45606     45705: gap of unknown length
            *      45706     64208: contig of 18503 bp in length
            *      64209     64308: gap of unknown length
            *      64309     70434: contig of 6126 bp in length
            *      70435     70534: gap of unknown length
            *      70535    123863: contig of 53329 bp in length
            *     123864    123963: gap of unknown length
            *     123964    133638: contig of 9675 bp in length
            *     133639    133738: gap of unknown length
            *     133739    151887: contig of 18149 bp in length
            *     151888    151987: gap of unknown length
            *     151988    161829: contig of 9842 bp in length
            *     161830    161929: gap of unknown length
            *     161930    181367: contig of 19438 bp in length
            *     181368    181467: gap of unknown length
            *     181468    207821: contig of 26354 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207821
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="RP42-405D15"
                     /clone_lib="RP42"
     misc_feature    1. .22661
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             22662. .22761
                     /estimated_length=unknown
     misc_feature    22762. .45605
                     /note="assembly_fragment"
     gap             45606. .45705
```

```
                     /estimated_length=unknown
     misc_feature    45706. .64208
                     /note="assembly_fragment"
     gap             64209. .64308
                     /estimated_length=unknown
     misc_feature    64309. .70434
                     /note="assembly_fragment"
     gap             70435. .70534
                     /estimated_length=unknown
     misc_feature    70535. .123863
                     /note="assembly_fragment"
     gap             123864. .123963
                     /estimated_length=unknown
     misc_feature    123964. .133638
                     /note="assembly_fragment"
     gap             133639. .133738
                     /estimated_length=unknown
     misc_feature    133739. .151887
                     /note="assembly_fragment"
     gap             151888. .151987
                     /estimated_length=unknown
     misc_feature    151988. .161829
                     /note="assembly_fragment"
     gap             161830. .161929
                     /estimated_length=unknown
     misc_feature    161930. .181367
                     /note="assembly_fragment"
     misc_feature    175822. .207821
                     /note="clone overlaps with GenBank Accession Number
                     AC120515 clone RP42-98H15 (center project name dje)"
     gap             181368. .181467
                     /estimated_length=unknown
     misc_feature    181468. .207821
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 14;  Length 207821;
  Best Local Similarity   53.7%;  Pred. No. 57;
  Matches   65;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          8 CTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGG 67
              || | |||     |  | |||  | | |    |||| |   || | |||  | |   |
Db      38349 CTCCCCGTTTGCCTCCCCTCGGCCCTCAGCTCCGACCAGAGCCTCCACCTGGCTCTCCCA 38408

Qy         68 GGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTTCAG 127
              || ||| |||||||   | ||  | ||  ||||||||  | | |  |||      | |||
Db      38409 GGCCTTGTTGTTGTCTAGTTGCTAAGTCATGTCCAACTCTGTGACCCCGTGGACTTGCAG 38468

Qy        128 C 128
              |
Db      38469 C 38469


RESULT 13
AF105113
LOCUS       AF105113                5832 bp    DNA     linear   BCT 09-SEP-1999
DEFINITION  Streptococcus pneumoniae type 19A putative oligosaccharide repeat
            unit transporter (cps19AJ) gene, partial cds; UDP-N-acetyl
            glucosamine-2-epimerase (cps19AK), glucose-1-phosphate thymidylyl
            transferase (cps19AL), dTDP-4-keto-6-deoxyglucose-3,5-epimerase
            (cps19AM), dTDP-glucose-4,6-dehydratase (cps19AN), and
            dTDP-L-rhamnose synthase (cps19AO) genes, complete cds; and AliA
            (aliA) gene, partial cds.
ACCESSION   AF105113
VERSION     AF105113.1  GI:4406246
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 5832)
```

```
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Comparative genetics of capsular polysaccharide biosynthesis in
            Streptococcus pneumoniae types belonging to serogroup 19
  JOURNAL   J. Bacteriol. 181 (17), 5355-5364 (1999)
   PUBMED   10464207
REFERENCE   2  (bases 1 to 5832)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1998) Molecular Microbiology Unit, Women's and
            Children's Hospital, 72 King William Rd, Adelaide, SA 5006,
            Australia
FEATURES             Location/Qualifiers
     source          1. .5832
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="19A2"
                     /serotype="19A"
                     /db_xref="taxon:1313"
                     /note="serogroup: 19"
     gene            <1. .818
                     /gene="cps19AJ"
     CDS             <1. .818
                     /gene="cps19AJ"
                     /note="Cps19AJ"
                     /codon_start=3
                     /transl_table=11
                     /product="putative oligosaccharide repeat unit
                     transporter"
                     /protein_id="AAD19911.1"
                     /db_xref="GI:4406247"
                     /translation="SKFIFNIKESFDNIRVGINLMLSNIASSLIIGIVRLGIQWNWNI
                     ETFGKVSLTLSVSNLLMTFINAIGLVVFPLLRRTKAENLPKIYSNLRNVLMLIMFAIL
                     LIYYPLKIVLDLWLPAYQDALIFMTLIFPMSVYEGKMALVINTYLKALRMERDILRVN
                     ALVMLISMGVTLVTTYLLNSLELTVVSIVVLLALRSIIAELILSKKLDVSVKKDIVLE
                     FLLTLVFISSSWYLPIGLAVIVYTIAYGLYLYLKHEDIKTYLAYFKASKKTSN"
     gene            883. .1971
                     /gene="cps19AK"
     CDS             883. .1971
                     /gene="cps19AK"
                     /note="Cps19AK"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-N-acetyl glucosamine-2-epimerase"
                     /protein_id="AAD19912.1"
                     /db_xref="GI:4406248"
                     /translation="MKKIMLVFGTRPEAIKMCPLVNELKKHEDMETIVCVTGQHKEMV
                     SPVLDLFGVVPDYDLEIMKANQTLFSITTSILEKIKPVLEKEQPDIVLIHGDTTTTYA
                     AALAAFYLGIKVGHVEAGLRTYNLQSPFPEEFNRQSTSIIATYHFAPTELAKENLLKE
                     GRENVYVTGNTVIDALTTTVQEDYTHTHLDLNANNRLILLTAHRRENLGEPMRHMFRA
                     VKRVLNEYEDVKVIYPIHKNPLVRETAAEIFGDTERIQIIEPLDVLDFHNFMNQSYMI
                     LTDSGGVQEEAPSLGKPVLVMRDTTERPEGVAAGTLKLVGTDEETIYQNFKLLLDDSG
                     EYKKMSQASNPYGNGDASQQIVQILRGI"
     gene            2134. .3003
                     /gene="cps19AL"
     CDS             2134. .3003
                     /gene="cps19AL"
                     /note="Cps19AL"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyl transferase"
                     /protein_id="AAD19913.1"
                     /db_xref="GI:4406249"
                     /translation="MKGIILAGGSGTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLA
                     GIKDILIISTPQDLPRFKDLLLDGSEFGIKLSYAEQPSPDGLAQAFLIGEEFIGDDSV
                     ALILGDNIYHGPGLSTMLQKAAKKEKGATVFGYQVKDPERFGVVEFDTDMNAISIEEK
                     PEYPRSNYAVTGLYFYDNDVVEIAKQIKPSARGELEITDVNKAYLNRGDLSVELMGRG
                     FAWLDTGTHESLLEASQYIETVQRMQNVQVANLEEISYRMGYISREDVLELAQPLKKN
                     EYGRYLLRLIGEA"
     gene            3004. .3597
                     /gene="cps19AM"
     CDS             3004. .3597
                     /gene="cps19AM"
                     /note="Cps19AM"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-4-keto-6-deoxyglucose-3,5-epimerase"
                     /protein_id="AAD19914.1"
                     /db_xref="GI:4406250"
                     /translation="MSDNFFGKTLVVRKIDAIPGLLEFDIPVHGDNRGWFKENFQKEK
                     MLPLGFPESFFAAGKLQNNVSFSRKNVLRGLHAEPWDKYISVADDGKVLGSWVDLREG
                     ETFGNVYQTEIDASKGIFVPRGVANGFQVLSDTVSYSYLVNDYWALELKPKYAFVNYA
                     DPALGIEWENLPEAEVSEADKHHPLLRDVKPLTKDEL"
     gene            3609. .4655
                     /gene="cps19AN"
     CDS             3609. .4655
                     /gene="cps19AN"
                     /note="Cps19AN"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-glucose-4,6-dehydratase"
                     /protein_id="AAD19915.1"
                     /db_xref="GI:4406251"
                     /translation="MTEYKKIIVTGGAGFIGSNFVHYVYNNFPDVHVTVLDKLTYAGN
                     RANIEEILGDRVELVVGDIADAALVDKLAAEADAIVHYAAESHNDNSLNDPSPFIHTN
                     FIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGHGEGLGEKFTAETKYNPSSP
                     YSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILSGIKPKLY
                     GEGKNVRDWIHTNDHSSGVWTILTKGQIGETYLIGADGEKNNKEVLELILKEMGQPAD
                     AYDHVTDRAGHDLRYAIDASKLRDELGWKPEFTNFEAGLKETIKWYTDNQDWWKSEKE
                     AVEANYAKTQQVIK"
     stem_loop       4659. .4704
     gene            complement(4717. .5571)
                     /gene="cps19AO"
     CDS             complement(4717. .5571)
                     /gene="cps19AO"
                     /note="Cps19AO"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-L-rhamnose synthase"
                     /protein_id="AAD19916.1"
                     /db_xref="GI:4406252"
                     /translation="MILITGANGQLGTELRYLLNERNVDYVAVDVTEMDITNSEMVEK
                     VFAEVKPTLVYHCAAYTAVDAAEDEGKELDFAINVTGTENVAKGSEKYGAILVYISTD
                     YVFDGKKPVGQEWEVDDIPDPQTEYGRTKRMGEELVENLTSQHYIIRTAWVFGNYGRN
                     FVFTMQSLAETHKTLTVVSDQYGRPTWTRTLAEFMTYVTENQKEYGYYHLSNDAAEDT
                     TWYDFASEILKESDVELLPVDSSKFPAKAKRPLNSTMSLAKAKATGFVIPTWQDALKE
                     FYKQEVRK"
     gene            5817. .>5832
                     /gene="aliA"
     CDS             5817. .>5832
                     /gene="aliA"
                     /codon_start=1
                     /transl_table=11
                     /product="AliA"
                     /protein_id="AAD19917.1"
                     /db_xref="GI:4406253"
                     /translation="MKSSK"
ORIGIN

  Query Match             24.0%;  Score 31;  DB 1;  Length 5832;
  Best Local Similarity   59.8%;  Pred. No. 79;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
               ||||| ||||  ||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        2722 ATTACAGACGTTAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 2781

Qy          92 CGTGTTGTCCAACCGTATTATTCCGGA 118
               |||| | |      ||   || | |||
Db        2782 CGTGGTTTTGCCTGGTTGGATACGGGA 2808


RESULT 14
CR931675
LOCUS       CR931675               18617 bp    DNA     linear   BCT 30-JUN-2005
DEFINITION  Streptococcus pneumoniae strain Nr. 141/68 (serotype 19a).
ACCESSION   CR931675
VERSION     CR931675.1  GI:68643359
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 18617)
  AUTHORS   Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
            Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
            Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
            Reeves,P., Parkhill,J. and Spratt,B.G.
  TITLE     Genetic analysis of the capsular biosynthetic locus from all 90
            serotypes of Streptococcus pneumoniae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 18617)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
            Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT     NOTE: This sequence was generated from a PCR product representing
            the region from dexB to aliA and is not necessarily responsible for
            the expressed capsule serotype. For a detailed description of how
            CDS products were predicted see the associated publication.
FEATURES             Location/Qualifiers
     source          1. .18617
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /strain="Nr. 141/68"
                     /db_xref="taxon:1313"
     gene            <1. .136
                     /gene="dexB"
                     /locus_tag="SPC19A_0001"
     CDS             <1. .136
                     /gene="dexB"
                     /locus_tag="SPC19A_0001"
                     /codon_start=2
                     /transl_table=11
                     /protein_id="CAI33621.1"
                     /db_xref="GI:68643360"
                     /translation="ANEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVEMTD
                     "
     gene            complement(267. .1697)
                     /gene="tnp"
                     /locus_tag="SPC19A_0002"
                     /pseudo
     CDS             complement(267. .1697)
                     /gene="tnp"
                     /locus_tag="SPC19A_0002"
                     /note="member of homology group 0017 25
                     member of homology group 0017 25"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
                     /product="putative IS1202 transposase"
     gene            1852. .3297
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
     CDS             1852. .3297
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
                     /note="member of homology group 0000 90"
                     /codon_start=1
                     /transl_table=11
                     /product="integral membrane regulatory protein Wzg"
                     /protein_id="CAI33623.1"
                     /db_xref="GI:68643361"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
                     RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSEIENVTQLMSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEHPDYASKIKKIYTKGFTKKVEAP
                     KTSKDQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKLIDLLGGVDVYNDQ
```

```
                     DFTSLHGKFHFPVGNVHLDSEQALGFVRERYSLADGDHDRGRNQQKVIAAILQKLTSS
                     EALKNYSRIIDSLQDSIQTNMPLETMINLVNAQLESGGTYKVNSQDLKGMGRTDLPSY
                     AMPDSNLYMMEINDSSLASVKTAIQDVLEGR"
     misc_feature    1852. .1968
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
                     /note="Signal peptide predicted for SPC0963 by SignalP 2.0
                     HMM (Signal peptide probability 0.999) with cleavage site
                     probability 0.540 between residues 39 and 40"
     misc_feature    join(1900. .1959,1987. .2055,2074. .2142)
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
                     /note="3 probable transmembrane helices predicted for
                     SPC0963 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
     misc_feature    2059. .2403
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
                     /note="HMMPfam hit to PF02916, DNA polymerase processivity
                     factor, score 6.7e-65"
     misc_feature    2581. .3024
                     /gene="wzg"
                     /locus_tag="SPC19A_0003"
                     /note="HMMPfam hit to PF03816, Cell envelope-related
                     transcriptional attenuator, score 2.8e-72"
     gene            3299. .4030
                     /gene="wzh"
                     /locus_tag="SPC19A_0004"
     CDS             3299. .4030
                     /gene="wzh"
                     /locus_tag="SPC19A_0004"
                     /EC_number="3.1.3.48"
                     /note="member of homology group 0001 90"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-tyrosine phosphatase Wzh"
                     /protein_id="CAI33624.1"
                     /db_xref="GI:68643362"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTSDVLDKLEKNRIPTLNNS
                     RYALIEFSMNTPYRDIHSALSKILMLGITPVIAHIERYDALENNEKRVRELINMGCYT
                     QVNSSHVLKSKLFGERYKFMKKRAQYFLERDLVHIIASDMHNVDSRPPHMAEAYDLVS
                     QKYGEAKAQELFIDNPRKIVMDQLI"
     misc_feature    3611. .3925
                     /gene="wzh"
                     /locus_tag="SPC19A_0004"
                     /note="HMMPfam hit to PF02811, PHP, C-terminal, score
                     1e-11"
     gene            4036. .4731
                     /gene="wzd"
                     /locus_tag="SPC19A_0005"
     CDS             4036. .4731
                     /gene="wzd"
                     /locus_tag="SPC19A_0005"
                     /note="member of homology group 0002 90"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide biosynthesis protein
                     Wzd"
                     /protein_id="CAI33625.1"
                     /db_xref="GI:68643363"
                     /translation="MMKEQNTIEIDVFQLLKTLWKHKLIILLVALVTGAGAFAYSTFI
                     VKPEYTSTTRIYVVNRNQGDKPGLTNQDLQAGTYLVKDYREIILSQDVLEKVATNLKL
                     DMPAKTLTSKVQVTVPADTRIVSISVKDKQPEEASRIANSLREVAAEKIIAVTRVSDV
                     TTLEEARPAMTPSSPNVRRNTLVGFLGGAAVTVITVLLIELLDTRVKRPEDVEDVLQM
                     PLLGVVPDFDKMK"
     misc_feature    4057. .4473
                     /gene="wzd"
                     /locus_tag="SPC19A_0005"
                     /note="HMMPfam hit to PF02706, Lipopolysaccharide
                     biosynthesis, score 1.6e-56"
     misc_feature    join(4105. .4173,4573. .4641)
                     /gene="wzd"
                     /locus_tag="SPC19A_0005"
                     /note="2 probable transmembrane helices predicted for
                     SPC0965 by TMHMM2.0 at aa 24-46 and 180-202"
     misc_feature    4579. .4656
                     /gene="wzd"
                     /locus_tag="SPC19A_0005"
                     /note="PS00217 Sugar transport proteins signature 2."
     gene            4741. .5430
                     /gene="wze"
                     /locus_tag="SPC19A_0006"
     CDS             4741. .5430
                     /gene="wze"
                     /locus_tag="SPC19A_0006"
                     /EC_number="2.7.1.112"
                     /note="member of homology group 0003 90"
                     /codon_start=1
                     /transl_table=11
                     /product="tyrosine-protein kinase Wze"
                     /protein_id="CAI33626.1"
                     /db_xref="GI:68643364"
                     /translation="MPTLEISQAKLELVKKAEEYYNALCTNLQLSGDDLKVFSITSVK
                     AGEGKTTTSTNIAWAFARAGYKTLLIDADMRNSVMSGVFKSRERITGLTEFLSGTTDL
                     SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFTTMLETLRKYFDYIVVDTAPVGVVI
                     DAAIIMQKCDASILVTKAGETKRRDLQKAKEQLEQTGKSCLGVVLNKLDTSVEKYGSY
                     GSYGNYRKQKK"
     gene            5445. .6806
                     /gene="wchA"
                     /locus_tag="SPC19A_0007"
     CDS             5445. .6806
                     /gene="wchA"
                     /locus_tag="SPC19A_0007"
                     /note="member of homology group 0005 66"
                     /codon_start=1
                     /transl_table=11
                     /product="undecaprenylphosphate glucosephosphotransferase
                     WchA (initial sugar transferase)"
                     /protein_id="CAI33627.1"
                     /db_xref="GI:68643365"
                     /translation="MSEKLAKPSVAIVQSFLVILLTYLLSAVRETEIVSTTAIVLYIL
                     HYFVFYISDYGRNFFKRGYLIELVQTLKYILFFALAISISNFFLEDRFSISRRGMIYF
                     LLLHVFLVYVLNPFIKWYWKRAYPNFKGSKKILLLTATSRVEKVLDRLIESDDVVGEL
                     VAVSVLDKPDFQHDDLKVVAEGEIVNFATHEVVDEVFINLPSEKYNIGELVSQFETMG
                     IDVTVNLNAFDWARNKQICEMAGLNVVTFSTTFYKTSHVIAKRIIDIVGSLVGLILCG
                     LVSIVLVPLIRKDGGSAFFAQTRIGKNGRHFTFYKFRSMCVDAEEKKRELMEQNTMQG
                     GMFKVDDDPRITKIGHFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPEQK
                     RRLSFKPGITGLWQVSGRSEIKNFDEVVKLDVAYIDGWTIWKDIEILLKTVKVVLMKD
                     GAK"
     misc_feature    join(5469. .5528,5541. .5594,5628. .5696,5739. .5807,
                     6237. .6305)
                     /gene="wchA"
                     /locus_tag="SPC19A_0007"
                     /note="5 probable transmembrane helices predicted for

  Query Match             24.0%;  Score 31;  DB 1;  Length 18617;
  Best Local Similarity   59.8%;  Pred. No. 78;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
               ||||| ||||  ||  |  |  ||| | || || | |  |||| ||||| | || || |
Db       13947 ATTACAGACGTTAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 14006

Qy          92 CGTGTTGTCCAACCGTATTATTCCGGA 118
               |||| | |      ||   || | |||
Db       14007 CGTGGTTTTGCCTGGTTGGATACGGGA 14033


RESULT 15
AF094575
LOCUS       AF094575               18754 bp    DNA     linear   BCT 09-SEP-1999
DEFINITION  Streptococcus pneumoniae serotype 19A DexB (dexB) gene, partial
            sequence; capsular polysaccharide biosynthesis operon, complete
            sequence; and oligopeptide permease AliA (aliA) gene, partial cds.
ACCESSION   AF094575
VERSION     AF094575.1  GI:3907597
```

```
KEYWORDS    .
SOURCE      Streptococcus pneumoniae
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Analysis of the 5' portion of the type 19A capsule locus identifies
            two classes of cpsC, cpsD, and cpsE genes in Streptococcus
            pneumoniae
  JOURNAL   J. Bacteriol. 181 (11), 3599-3605 (1999)
   PUBMED   10348877
REFERENCE   2  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Comparative genetics of capsular polysaccharide biosynthesis in
            Streptococcus pneumoniae types belonging to serogroup 19
  JOURNAL   J. Bacteriol. 181 (17), 5355-5364 (1999)
   PUBMED   10464207
REFERENCE   3  (bases 1 to 18754)
  AUTHORS   Morona,J.K., Morona,R. and Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1998) Microbial Microbiology Unit, Women's and
            Children's Hospital, King William Rd, North Adelaide, SA 5006,
            Australia
FEATURES             Location/Qualifiers
     source          1. .18754
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /serotype="19A"
                     /db_xref="taxon:1313"
     gene            <1. .9
                     /gene="dexB"
     CDS             <1. .9
                     /gene="dexB"
                     /codon_start=1
                     /transl_table=11
                     /product="DexB"
                     /protein_id="AAC78662.1"
                     /db_xref="GI:3907598"
                     /translation="VE"
     repeat_region   complement(161. .1647)
                     /note="insertion sequence; 3' truncated"
                     /rpt_family="IS1202"
                     /rpt_type=dispersed
     gene            1733. .16682
                     /gene="capsular polysaccharide biosynthesis operon"
     gene            1733. .3178
                     /gene="cps19aA"
     CDS             1733. .3178
                     /gene="cps19aA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative regulatory protein Cps19aA"
                     /protein_id="AAC78663.1"
                     /db_xref="GI:3907599"
                     /translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFTYNILAF
                     RYLNLVVPALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
                     LNATSNYSEYSISVAVLADSEIENVTQLMSVTAPTGTDNENIQKLLADIKSSQNTDLT
                     VNQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEHPDYASKIKKIYTKGFTKKVEAP
                     KTSKDQSFNIYVSGIDTYGPISSVSRSYVNILMTVNRDTKKILLTTTPRDAYVPIADG
                     GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKLIDLLGGVDVYNDQ
                     DFTSLHGKFHFPVGNVHLDSEQALGFVRERYSLADGDHDRGRNQQKVIAAILQKLTSS
                     EALKNYSRIIDSLQDSIQTNMPLETMINLVNAQLESGGTYKVNSQDLKGMGRTDLPSY
                     AMPDSNLYMMEINDSSLASVKTAIQDVLEGR"
     gene            3180. .3911
                     /gene="cps19aB"
     CDS             3180. .3911
                     /gene="cps19aB"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aB"
                     /protein_id="AAC78664.1"
                     /db_xref="GI:3907600"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVASDLVIAYGAEIYYTSDVLDKLEKNRIPTLNNS
                     RYALIEFSMNTPYRDIHSALSKILMLGITPVIAHIERYDALENNEKRVRELINMGCYT
                     QVNSSHVLKSKLFGERYKFMKKRAQYFLERDLVHIIASDMHNVDSRPPHMAEAYDLVS
                     QKYGEAKAQELFIDNPRKIVMDQLI"
     gene            3917. .4612
                     /gene="cps19aC"
     CDS             3917. .4612
                     /gene="cps19aC"
                     /function="chain length regulation and export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aC"
                     /protein_id="AAC78665.1"
                     /db_xref="GI:3907601"
                     /translation="MMKEQNTIEIDVFQLLKTLWKHKLIILLVALVTGAGAFAYSTFI
                     VKPEYTSTTRIYVVNRNQGDKPGLTNQDLQAGTYLVKDYREIILSQDVLEKVATNLKL
                     DMPAKTLTSKVQVTVPADTRIVSISVKDKQPEEASRIANSLREVAAEKIIAVTRVSDV
                     TTLEEARPAMTPSSPNVRRNTLVGFLGGAAVTVITVLLIELLDTRVKRPEDVEDVLQM
                     PLLGVVPDFDKMK"
     gene            4622. .5311
                     /gene="cps19aD"
     CDS             4622. .5311
                     /gene="cps19aD"
                     /function="chain length regulation and export"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aD"
                     /protein_id="AAC78666.1"
                     /db_xref="GI:3907602"
                     /translation="MPTLEISQAKLELVKKAEEYYNALCTNLQLSGDDLKVFSITSVK
                     AGEGKTTTSTNIAWAFARAGYKTLLIDADMRNSVMSGVFKSRERITGLTEFLSGTTDL
                     SQGLCDTNIENLFVIQAGSVSPNPTALLQSKNFTTMLETLRKYFDYIVVDTAPVGVVI
                     DAAIIMQKCDASILVTKAGETKRRDLQKAKEQLEQTGKSCLGVVLNKLDTSVEKYGSY
                     GSYGNYRKQKK"
     gene            5326. .6687
                     /gene="cps19aE"
     CDS             5326. .6687
                     /gene="cps19aE"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyl-1-phosphate transferase Cps19aE"
                     /protein_id="AAC78667.1"
                     /db_xref="GI:3907603"
                     /translation="MSEKLAKPSVAIVQSFLVILLTYLLSAVRETEIVSTTAIVLYIL
                     HYFVFYISDYGRNFFKRGYLIELVQTLKYILFFALAISISNFFLEDRFSISRRGMIYF
                     LLLHVFLVYVLNPFIKWYWKRAYPNFKGSKKILLLTATSRVEKVLDRLIESDDVVGEL
                     VAVSVLDKPDFQHDDLKVVAEGEIVNFATHEVVDEVFINLPSEKYNIGELVSQFETMG
                     IDVTVNLNAFDWARNKQICEMAGLNVVTFSTTFYKTSHVIAKRIIDIVGSLVGLILCG
                     LVSIVLVPLIRKDGGSAFFAQTRIGKNGRHFTFYKFRSMCVDAEEKKRELMEQNTMQG
                     GMFKVDDDPRITKIGHFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPEQK
                     RRLSFKPGITGLWQVSGRSEIKNFDEVVKLDVAYIDGWTIWKDIEILLKTVKVVLMKD
                     GAK"
     gene            6693. .7436
                     /gene="cps19aF"
     CDS             6693. .7436
                     /gene="cps19aF"
                     /codon_start=1
                     /transl_table=11
                     /product="putative N-acetyl-mannosamine transferase
                     Cps19aF"
                     /protein_id="AAC78668.1"
                     /db_xref="GI:3907604"
                     /translation="MNERIQILGITIDPLTMKETVDAVEQYVLKKHPLHLMGVNADKI
                     NQCYEDEKIKKIVNESGIINADGASVVLASKFLGTPVPERVAGIDLMKHLLELSNEKG
                     YSVYFFGAKEDVLQDMLKIFEERYPALNVVGYRNGYFSPEDEKHIQEDIKEKKPDFVF
                     VGITSPKKEYIIQSFMDNGINAVFMGVGGSFDVLSGHIKRAPLWMQNAHLEWLFRVAN
                     EPKRLFKRYFVGNVSFIRKVLRAKRGYKL"
     gene            7475. .8257
                     /gene="cps19aG"
     CDS             7475. .8257
                     /gene="cps19aG"
                     /function="putative protein involved in formation of
```

```
                     phospho-diester linkage"
                     /codon_start=1
                     /transl_table=11
                     /product="Cps19aG"
                     /protein_id="AAC78669.1"
                     /db_xref="GI:3907605"
                     /translation="MDAIKEFQKICKENNIDFFLRGGSVLGAVKYDGFIPWDDDMDIA
                     VPREGYDKLPGIFKDRIIAGKYQVLAYQYCDTLHCYFPRLFLLEDERKRLGLPRNTNL
                     GLHLIDIIPLDGAPNHSFLRKLYFGKVYWYRFLASLGTTYVGDHVDMHSAKQKLIIGF
                     FKKLGFAKLFPQNYVYRRLDNLYKKYDWKKQKYAGTINASLFAKEVMPVEIWGEGVEK
                     PFEDAFFKVPTEYDRYLKRLYGENYLHEEPSDDEKKSHLGGQ"
     gene            8259. .9137
                     /gene="cps19aH"
     CDS             8259. .9137
                     /gene="cps19aH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative rhamnosyl transferase Cps19aH"
                     /protein_id="AAC78670.1"
                     /db_xref="GI:3907606"
                     /translation="MFCYIILHYKVLEETISCVKSIKEGNSNTKQIVIIDNFSNNGTG
                     EKLQELYESDSEIDVLINHENAGFARGNNVAYQFAKEKYKPDFMVIMNNDIEIETEEF
                     EKIVTDIYRKEKFHLLGPDIFSTTYQLHQNPKRLTHYTYEEVKALNEKFKKGSQVSLA
                     LKIKCWLKSSKVLRTAIYQNRRKKKSVDYRKQVENPILHGSFIVYSRDLIEKEEYAFN
                     PNTFFYYETEILDYEAELKGYKRIYTPKIKVLHHQNVATNQVYTNLVEKTLFSNKCNF
                     ESTSYFLKLMKENEGV"
     gene            9138. .10472
                     /gene="cps19aI"
     CDS             9138. .10472
                     /gene="cps19aI"
                     /codon_start=1
                     /transl_table=11
                     /product="polysaccharide polymerase Cps19aI"
                     /protein_id="AAC78671.1"
                     /db_xref="GI:3907607"
                     /translation="MTYLFLLCLTLFLLTFFYFFAFNQDLIAPPVVMSVMFLISSVFA
                     LINVQNWNIEYSGLAYLLIISGIIVFSMPLLALNSPSLNTKIKVTDRLIDIQFWKIAL
                     TIIVDLIILYLYRREIHNLALSHGYTGSNFQWFFRNATSYEGELTVRTSIRVLIRIID
                     VSAYIFGYTFINNFFIYSHKRSKDLLLLVPFLIFISKTLLSGGRLDIIKILIAYVVMA
                     YIQQKRKVGWDKVISHKYMRLGFVGLIAGIPTFYYSLFLSGRSTTRTVFESISTYLGG
                     SIQHFNQYIQNPIGVAEVFGDESFVAIMNILGNLGFVNYNSTVHLEFRQLGITMGNVY
                     TFFRRPWHDFGLVGMYIFSFVVGVFFAIFYLKLRKSRAGFKLDIHTIIYSYFFYWIFL
                     SSIEQYSFTTISLFTLVFIVLVYLMAFFYWNLDFHRGKLVIKLSDTSIKSEN"
     gene            10497. .11924
                     /gene="cps19aJ"
     CDS             10497. .11924
                     /gene="cps19aJ"
                     /codon_start=1

  Query Match             24.0%;  Score 31;  DB 1;  Length 18754;
  Best Local Similarity   59.8%;  Pred. No. 78;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||||  ||  |  |  ||| | || || | |  |||| ||||| | || || |
Db      13828 ATTACAGACGTTAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 13887

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db      13888 CGTGGTTTTGCCTGGTTGGATACGGGA 13914

Search completed: March  7, 2006, 12:15:11
Job time : 1202.77 secs
```

```
                        GenCore version 5.1.7
               Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:          March  7, 2006, 10:55:04 ; Search time 304.617 Seconds
                                            (without alignments)
                                            2822.383 Million cell updates/sec

Title:           US-10-782-899-1_COPY_472_600
Perfect score:   129
Sequence:        1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters:      9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :        N_Geneseq_21:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002as:*
                 7:  geneseqn2002bs:*
                 8:  geneseqn2003as:*
                 9:  geneseqn2003bs:*
                 10:  geneseqn2003cs:*
                 11:  geneseqn2003ds:*
                 12:  geneseqn2004as:*
                 13:  geneseqn2004bs:*
                 14:  geneseqn2005s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                      %
Result               Query
  No.    Score   Match Length DB  ID                           Description
-------------------------------------------------------------------------------
     1      129  100.0    600  6  ABK52324                     Abk52324 DNA encod
c    2      129  100.0 156638  6  ABQ81850                     Abq81850 Bifidobac
     3     30.6   23.7  26309  4  AAS59535                     Aas59535 Propionib
     4     30.6   23.7  26309  8  ACF64464                     Acf64464 Propionib
     5     29.6   22.9   2453 10   ADE63675                     Ade63675 Human gen
     6     29.4   22.8    870 13   ADK44809                     Adk44809 Streptoco
     7     29.4   22.8    885 13   ADR91671                     Adr91671 Novel S.
     8     29.4   22.8    885 14   AEA55541                     Aea55541 Streptoco
c    9     29.4   22.8   1377 12   ADQ89190                     Adq89190 Non-natur
    10     29.4   22.8   3520  5  ADL45225                     Adl45225 Human ova
c   11     29.2   22.6    634  6  ABL01350                     Abl01350 Murine ap
c   12     29.2   22.6   1659 14   ADW69564                     Adw69564 DNA encod
    13     29.2   22.6   1760 11   AEA79104                     Aea79104 Human apo
    14     29.2   22.6   1800 13   ADR89493                     Adr89493 Apoptosis
    15     29.2   22.6   1836  6  ABL01352                     Abl01352 Murine ap
    16       29   22.5 110000 14   AEA61160_0                   Aea61160 Human NED
c   17     28.6   22.2   1146 11   ABD03998                     Abd03998 Pseudomon
    18     28.6   22.2   2184 11   ABD03637                     Abd03637 Pseudomon
    19     28.6   22.2   2868 11   ABD03707                     Abd03707 Pseudomon
```

```
    20     28.4   22.0    936  2  AAT79329                     Aat79329 DNA encod
    21     28.4   22.0    936  8  ABX11507                     Abx11507 DNA encod
    22     28.4   22.0    936 10   AAD58914                     Aad58914 Archaeogl
    23     28.4   22.0   2451  6  ABZ12259                     Abz12259 Arabidops
c   24     28.2   21.9    599 12   ACH79902                     Ach79902 Human gen
    25     28.2   21.9   1152  5  AAS73582                     Aas73582 DNA encod
    26     28.2   21.9  13540  8  ABZ74138                     Abz74138 Secreted
    27     28.2   21.9  13540  8  ADA98673                     Ada98673 Human sec
    28     28.2   21.9  13540 10   ABT16951                     Abt16951 Human sec
    29     28.2   21.9  13540 10   ABZ67695                     Abz67695 Human sec
c   30     28.2   21.9  18038  5  ADM20219                     Adm20219 Alternati
    31     28.2   21.9  18038  8  ABZ74137                     Abz74137 Secreted
    32     28.2   21.9  18038  8  ADA98672                     Ada98672 Human sec
    33     28.2   21.9  18038 10   ABT16950                     Abt16950 Human sec
    34     28.2   21.9  18038 10   ABZ67694                     Abz67694 Human sec
c   35       28   21.7 165221 11   ACN44524                     Acn44524 Mouse gen
c   36       28   21.7 165799 14   ADZ13004                     Adz13004 Murine ca
c   37     27.8   21.6   2037  4  AAS46220                     Aas46220 Human DNA
c   38     27.8   21.6   2037  8  ACA89670                     Aca89670 cDNA enco
c   39     27.8   21.6   2037  8  ACA73680                     Aca73680 Human sec
c   40     27.8   21.6   2037  8  ACA05995                     Aca05995 Human sec
c   41     27.8   21.6   2037  8  ACA66829                     Aca66829 cDNA enco
c   42     27.8   21.6   2037  8  ACF20404                     Acf20404 Human sec
c   43     27.8   21.6   2037  8  ACF19790                     Acf19790 Human sec
c   44     27.8   21.6   2037  8  ACD22078                     Acd22078 Human sec
c   45     27.8   21.6   2037  8  ACF13243                     Acf13243 Human sec


                                 ALIGNMENTS


RESULT 1
ABK52324
ID   ABK52324 standard; DNA; 600 BP.
XX
AC   ABK52324;
XX
DT   13-AUG-2002  (first entry)
XX
DE   DNA encoding cancer gene therapy associated protein.
XX
KW   Solid cancer; anaerobic bacteria; antitumour active protein; cancer;
KW   gene therapy; gene; ds.
XX
OS   Bifidobacterium longum.
XX
FH   Key             Location/Qualifiers
FT   CDS             193. .474
FT                   /*tag=  a
FT                   /product= "Cancer gene therapy protein"
XX
PN   JP2002097144-A.
XX
PD   02-APR-2002.
XX
PF   21-SEP-2001; 2001JP-00290187.
XX
PR   21-SEP-2000; 2000JP-00287688.
XX
PA   (AMAN/) AMANO A.
PA   (FUJI/) FUJIMORI M.
XX
DR   WPI; 2002-448201/48.
DR   P-PSDB; AAU96807.
XX
PT   Solid cancer therapy with anaerobic bacteria of Bifidobacterium sp. by
PT   tumor tissue specific delivery of a DNA encoding for an antitumor active
PT   protein or its precursor.
XX
PS   Claim 10; Page 16; 21pp; Japanese.
XX
CC   The invention describes a method of treating a solid cancer with
CC   anaerobic bacteria by site specific delivery of DNA encoding an
```

```
CC   antitumour active protein or its precursor. This sequence encodes a
CC   cancer gene therapy associated protein
XX
SQ   Sequence 600 BP; 132 A; 165 C; 161 G; 142 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 129;  DB 6;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 6e-34;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
ABQ81850/c
ID   ABQ81850 standard; DNA; 156638 BP.
XX
AC   ABQ81850;
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1106.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
OS   Synthetic.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-00102050.
XX
PR   30-JAN-2001; 2001EP-00102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT   a probe or primer for detecting and/or identifying Bifidobacterium longum
PT   in a biological sample.
XX
PS   Disclosure; SEQ ID NO 1106; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC   fusion protein, comprising a sequence selected from 1097 sequences given
CC   in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC   heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC   activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC   a probe) is useful for the detection and/or identification of
CC   Bifidobacterium longum in a biological sample. A carrier containing the
CC   lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
CC   used for preventing and/or treating diarrhoea brought about by pathogenic
CC   bacteria and/or rotavirus. The carrier is a food composition selected
CC   from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC   products, ice-creams, fermented cereal based products, milk based
CC   powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification. N.B. The sequence data for this patent is not represented
CC   in the printed specification but is based on sequence information
CC   supplied by the European Patent Office
XX
SQ   Sequence 156638 BP; 32098 A; 46491 C; 46415 G; 31634 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 129;  DB 6;  Length 156638;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-33;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141691 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 141632

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141631 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 141572

Qy        121 AGTTCAGCG 129
              |||||||||
Db     141571 AGTTCAGCG 141563


RESULT 3
AAS59535
ID   AAS59535 standard; DNA; 26309 BP.
XX
AC   AAS59535;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein encoding DNA #30.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant; ds.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US012865.
XX
PR   21-APR-2000; 2000US-0199047P.
PR   02-JUN-2000; 2000US-0208841P.
PR   07-JUL-2000; 2000US-0216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris.
XX
PS   Claim 1; SEQ ID NO 30; 1069pp; English.
XX
CC   Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC   Propionibacterium acnes immunogenic polypeptides. The proteins and their
```

```
CC   associated DNA sequences are used in the treatment, prevention and
CC   diagnosis of medical conditions caused by P. acnes. The disorders include
CC   SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC   osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
CC   infections of bone, joints and the central nervous system, however it is
CC   particularly involved in the inflammatory lesions associated with acne
CC   vulgaris. A method for detecting the presence or absence of P. acnes in a
CC   patient comprises contacting a sample with a binding agent that binds to
CC   the proteins of the invention and determining the amount of bound protein
CC   in the sample. The polypeptides may be used as antigens in the production
CC   of antibodies specific for P. acnes proteins. These antibodies can be
CC   used to downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC   polypeptides shown in AAU46704-AAU46985 and AAU67509. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;

  Query Match             23.7%;  Score 30.6;  DB 4;  Length 26309;
  Best Local Similarity   58.1%;  Pred. No. 11;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         37 TGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              ||||||| |  | ||||||    ||  |  |||||  || | |||   |||||||   ||
Db       9538 TGACGACGACCATCCCGACTCCCATATTGAGGGTCGCTTCGATGTCAGGCTGTGAGACGT 9597

Qy         97 TGTCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              || ||| | |    |    |    || || |||
Db       9598 TGCCCACCTGCTGGACGAGGTTGAAGATCGGCG 9630


RESULT 4
ACF64464
ID   ACF64464 standard; DNA; 26309 BP.
XX
AC   ACF64464;
XX
DT   17-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes DNA contig sequence #30.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine; ds.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Claim 1; SEQ ID NO 30; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
```

```
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a P. acnes DNA contig which is specifically claimed
CC   in the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;

  Query Match             23.7%;  Score 30.6;  DB 8;  Length 26309;
  Best Local Similarity   58.1%;  Pred. No. 11;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         37 TGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGT 96
              ||||||| |  | ||||||    ||  |  |||||  || | |||   |||||||   ||
Db       9538 TGACGACGACCATCCCGACTCCCATATTGAGGGTCGCTTCGATGTCAGGCTGTGAGACGT 9597

Qy         97 TGTCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              || ||| | |    |    |    || || |||
Db       9598 TGCCCACCTGCTGGACGAGGTTGAAGATCGGCG 9630


RESULT 5
ADE63675
ID   ADE63675 standard; DNA; 2453 BP.
XX
AC   ADE63675;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human gene NM_004046, SEQ ID NO 9619.
XX
KW   Human; ds; gene; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
```

```
DR   GENBANK; NM_004046.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC   specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2453 BP; 622 A; 536 C; 635 G; 660 T; 0 U; 0 Other;

  Query Match             22.9%;  Score 29.6;  DB 10;  Length 2453;
  Best Local Similarity   57.6%;  Pred. No. 12;
  Matches   53;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || ||    || |  |||| ||| || ||  | || | ||||||   ||| ||   ||
Db       1864 CCATCACTGCGGACAGATTTCTTGGAACAGAATTGTCTACAAAGGATCCGCCCTCAATTA 1923

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
               || ||||| |    | | || | |   ||||
Db       1924 ACGTGGTCTGTCGTATCTCGTTCGGATCCGTG 1955


RESULT 6
ADK44809
ID   ADK44809 standard; DNA; 870 BP.
XX
AC   ADK44809;
XX
DT   24-FEB-2005  (first entry)
XX
DE   Streptococcus pneumoniae gene, Seq ID No 1324.
XX
KW   ds; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS   Streptococcus pneumoniae.
XX
PN   US6699703-B1.
XX
PD   02-MAR-2004.
XX
PF   26-MAY-2000; 2000US-00583110.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
```

```
PR   30-JUN-1998;   98US-00107433.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm L,  Bush D,  Zeng Q,  Opperman T,  Houseweart CE;
XX
DR   WPI; 2004-212399/20.
DR   P-PSDB; ADK47470.
XX
PT   New nucleic acid molecules and polypeptides useful for diagnosing,
PT   preventing and treating pathological conditions resulting from bacterial
PT   infection, e.g. Streptococcus pneumoniae infection, and in drug
PT   screening.
XX
PS   Disclosure; SEQ ID NO 1324; 301pp; English.
XX
CC   The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC   and polypeptides. The nucleic acids and proteins are useful for
CC   diagnosing, preventing and treating pathological conditions resulting
CC   from bacterial infection, such as S. pneumoniae infection. These may also
CC   be used for drug screening procedures. The present sequence represents a
CC   Streptococcus pneumoniae nucleic acid of the invention. Note: The
CC   sequence data for this patent did not appear in the printed specification
CC   but was obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 870 BP; 237 A; 166 C; 215 G; 252 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 13;  Length 870;
  Best Local Similarity   58.6%;  Pred. No. 10;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        589 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 648

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        649 CGTGGTTTTGCCTGGTTGGATACGGGA 675


RESULT 7
ADR91671
ID   ADR91671 standard; DNA; 885 BP.
XX
AC   ADR91671;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Novel S. pneumoniae DNA sequence, SEQ ID 306.
XX
KW   Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW   bacterial infection.
XX
OS   Streptococcus pneumoniae.
XX
PN   US6800744-B1.
XX
PD   05-OCT-2004.
XX
PF   30-JUN-1998;   98US-00107433.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA,  Bush D;
XX
DR   WPI; 2004-697205/68.
DR   P-PSDB; ADR94274.
XX
PT   New isolated nucleic acid encoding a Streptococcus pneumoniae
```

```
PT   polypeptide, useful for diagnosing, preventing and/or treating
PT   pathological conditions resulting from the bacterial infection.
XX
PS   Disclosure; SEQ ID NO 306; 151pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising a sequence
CC   encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC   fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC   ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC   ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC   ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC   ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC   sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC   hybridisable under high stringency conditions to the nucleotide sequence.
CC   The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC   Also included are a recombinant expression vector comprising the isolated
CC   nucleic acid cited above operably linked to a transcription regulatory
CC   element, a cell comprising the recombinant expression vector and a probe
CC   comprising at least 20 consecutive nucleotides of the nucleotide
CC   sequences as cited above. The methods and compositions of the present
CC   invention are useful for the diagnosis, prevention and/or treatment of
CC   pathological conditions resulting from bacterial infection by
CC   Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC   otitis media. The present sequence is one of the 2603 disclosed S.
CC   pneumoniae nucleic acid sequences. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ   Sequence 885 BP; 243 A; 168 C; 218 G; 256 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 13;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 10;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        664 CGTGGTTTTGCCTGGTTGGATACGGGA 690


RESULT 8
AEA55541
ID   AEA55541 standard; DNA; 885 BP.
XX
AC   AEA55541;
XX
DT   25-AUG-2005  (first entry)
XX
DE   Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:306.
XX
KW   bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW   vaccine; gene; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   US2005136404-A1.
XX
PD   23-JUN-2005.
XX
PF   10-JUL-2003; 2003US-00617320.
XX
PR   02-JUL-1997;   97US-0051553P.
PR   12-MAY-1998;   98US-0085131P.
PR   30-JUN-1998;   98US-00107433.
XX
PA   (DOUC/) DOUCETTE-STAMM L A.
PA   (BUSH/) BUSH D.
XX
PI   Doucette-Stamm LA,  Bush D;
```

```
XX
DR   WPI; 2005-477576/48.
DR   P-PSDB; AEA58144.
XX
PT   New isolated nucleic acid molecules and encoded polypeptides useful for
PT   diagnosing, preventing or treating bacterial infections, particularly
PT   Streptococcus pneumoniae infection.
XX
PS   Claim 1; SEQ ID NO 306; 144pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule for detecting,
CC   preventing or treating pathological conditions resulting from bacterial
CC   infection. The isolated nucleic acid comprises: (a) any of the 2603
CC   nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
CC   encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC   2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC   sequence of at least 8 nucleotides in length, where the sequence is
CC   hybridizable to a nucleic acid having any of the nucleotide sequences in
CC   (a). Also described: (1) a recombinant expression vector comprising the
CC   above nucleic acid operably linked to a transcription regulatory element;
CC   (2) a cell comprising the recombinant expression vector; (3) producing an
CC   S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC   consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
CC   treating a subject for S. pneumoniae infection; (6) a recombinant or
CC   substantially pure preparation of an S. pneumoniae polypeptide or its
CC   fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC   (7) a vaccine composition for preventing or treating an S. pneumoniae
CC   infection, comprising an amount of the above nucleic acid or polypeptide;
CC   (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC   (9) a computer readable medium having recorded the nucleotide sequences
CC   of AEA55236 to AEA57838; (10) a computer based system for identifying
CC   fragments of the Streptococcus genome of commercial importance. The
CC   composition and methods are useful for diagnosing, preventing or treating
CC   bacterial infections, particularly S. pneumoniae infection. The present
CC   sequence represents a S. pneumoniae ORF nucleic acid sequence from the
CC   present invention. Note - The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from the USPTO web site.
XX
SQ   Sequence 885 BP; 243 A; 168 C; 218 G; 256 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 14;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 10;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        664 CGTGGTTTTGCCTGGTTGGATACGGGA 690


RESULT 9
ADQ89190/c
ID   ADQ89190 standard; DNA; 1377 BP.
XX
AC   ADQ89190;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Non-natural ND4 mitchondrial protein coding sequence.
XX
KW   gene therapy; ND4 mitochondrial protein; ND4; cellular dysfunction;
KW   mtDNA mutation; Leber Hereditary Optic Neuropathy;
KW   mitochondrial gene mutation; human; gene; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2004142419-A1.
XX
```

```
PD   22-JUL-2004.
XX
PF   17-OCT-2003; 2003US-00687677.
XX
PR   18-OCT-2002; 2002US-0419435P.
XX
PA   (GUYJ/) GUY J.
XX
PI   Guy J;
XX
DR   WPI; 2004-579908/56.
XX
PT   New non-naturally occurring nucleic acid comprises a nucleotide sequence
PT   that encodes a functional ND4 mitochondrial protein, useful for reducing
PT   cellular dysfunction caused by mitochondrial gene mutations.
XX
PS   Claim 7; SEQ ID NO 1; 16pp; English.
XX
CC   The invention describes a non-naturally occurring nucleic acid comprising
CC   a nucleotide sequence that encodes a functional ND4 mitochondrial protein
CC   and differs from a naturally occurring nucleic acid that encodes a ND4
CC   mitochondrial protein by at least one codon substitution. Also described
CC   are: a cell into which has been introduced the non-naturally occurring
CC   nucleic acid above; and reducing dysfunction in a cell caused by a mtDNA
CC   mutation associated with Leber Hereditary Optic Neuropathy. Specifically
CC   claimed is non-naturally occurring ND4 nucleic acid comprising 1377 base
CC   pairs (SEQ ID NO. 1), fully defined in the specification. The nucleic
CC   acid is useful for reducing cellular dysfunction caused by mitochondrial
CC   gene mutations. Compositions comprising the non-naturally occurring
CC   nucleic acids are also useful for treating mtDNA mutations in animal
CC   subjects, including humans. This sequence represents a non-naturally
CC   occurring ND4 mitochondrial protein encoding polynucleotide sequence.
XX
SQ   Sequence 1377 BP; 330 A; 436 C; 341 G; 270 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 12;  Length 1377;
  Best Local Similarity   63.4%;  Pred. No. 12;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         57 GAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
              | || | | ||  || | ||||||  ||| |||| |  | || || || || || ||  |
Db       1292 GTGAAGCTGGGCTTCATGTTGTTGATGTGGTGTGTCAGGCTGCCCCACTGTGTTGTTGTG 1233

Qy        117 GACTAGTTCAG 127
               ||  || |||
Db       1232 AACATGTACAG 1222


RESULT 10
ADL45225
ID   ADL45225 standard; DNA; 3520 BP.
XX
AC   ADL45225;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human ovarian cancer DNA marker #19115.
XX
KW   Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS   Homo sapiens.
XX
PN   WO200170979-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US009126.
XX
PR   21-MAR-2000; 2000US-0191031P.
PR   25-MAY-2000; 2000US-0207124P.
PR   15-JUN-2000; 2000US-0211940P.
PR   07-JUL-2000; 2000US-0216820P.
PR   25-JUL-2000; 2000US-0220661P.
```

```
PR   21-DEC-2000; 2000US-0257672P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Lillie J;
XX
DR   WPI; 2001-611502/70.
XX
PT   Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT   cancer cells as compared to their normal non-cancerous ovarian cells are
PT   used to characterize stage, grade, histological type of ovarian cancer.
XX
PS   Disclosure; SEQ ID NO 19115; 106pp; English.
XX
CC   The invention relates to nucleic acid markers which are overexpressed in
CC   ovarian cancer cells as compared to their expression in normal (i.e. non-
CC   cancerous) ovarian cells. The invention also relates to polypeptides
CC   encoded by the markers, antibodies that selectively bind to the
CC   polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC   of developing ovarian cancer involving inhibiting expression of a gene
CC   corresponding to a marker of the invention and a method of treating a
CC   patient afflicted with ovarian cancer comprising providing to cells of
CC   the patient an antisense oligonucleotide complementary to a marker of the
CC   invention. The markers are useful for assessing if a patient is afflicted
CC   with ovarian cancer, which involves comparing the level of expression of
CC   a marker in a patient sample and a normal level of expression of the
CC   marker in a control non-ovarian cancer sample. A difference between the
CC   expression levels indicates ovarian cancer. The level of expression of a
CC   marker corresponds to a secreted protein or to a transcribed
CC   polynucleotide or its portion. The level of expression of the marker is
CC   assessed by detecting the presence in the sample, a protein or protein
CC   fragment corresponding to the marker. The presence of protein or protein
CC   fragment is detected using an antibody that specifically binds with the
CC   protein or protein fragment. Alternatively, the level of expression of
CC   the marker is assessed by detecting the presence of a transcribed
CC   polynucleotide which anneals with the marker or anneals with a portion of
CC   the polynucleotide comprising the marker, under stringent conditions. The
CC   marker is also used for monitoring the progression of ovarian cancer in a
CC   patient which involves detecting expression of the marker in a patient
CC   sample at a first point in time, repeating the method at a subsequent
CC   time and comparing the level of expression. The method is carried out
CC   using an ovarian tissue sample. A composition comprising a marker,
CC   polypeptide or antibody of the invention is used to treat ovarian cancer.
CC   This sequence represents a human ovarian cancer DNA marker of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3520 BP; 679 A; 939 C; 1047 G; 855 T; 0 U; 0 Other;

  Query Match             22.8%;  Score 29.4;  DB 5;  Length 3520;
  Best Local Similarity   54.1%;  Pred. No. 16;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || ||| ||  |      | |||      ||  | |    |   | |||   |  | |||
Db         27 CCCTCTACTGCTCATCTCTGCTTACCTTGTTTTTTAGTGGATTAATCCCTCACCTGGTGG 86

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTC 114
               ||  ||  |||| |||||| ||||| || || || || ||   || || |
Db         87 GCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTCATAATCC 137


RESULT 11
ABL01350/c
ID   ABL01350 standard; DNA; 634 BP.
XX
AC   ABL01350;
XX
DT   15-MAR-2002  (first entry)
XX
DE   Murine apoptosis related DNA sequence #15.
XX
```

```
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
XX
PF   30-MAY-2001; 2001DE-01026344.
XX
PR   14-JUL-2000; 2000DE-01034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides, useful
PT   for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration.
XX
PS   Claim 1; Page 137; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention
XX
SQ   Sequence 634 BP; 157 A; 184 C; 160 G; 133 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 6;  Length 634;
  Best Local Similarity   59.8%;  Pred. No. 11;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        245 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 186

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        185 ATCAATGCCTACTTGGTGGACG 164


RESULT 12
ADW69564/c
ID   ADW69564 standard; DNA; 1659 BP.
XX
AC   ADW69564;
XX
DT   24-MAR-2005  (first entry)
XX
DE   DNA encoding HIV-1 gp120 protein, U-323c11.
XX
KW   immunogenicity; anti-hiv; vaccine; immune stimulation; HIV-1 infection;
KW   gp120; gene; ds; U-323c11.
XX
OS   Human immunodeficiency virus 1.
XX
```

```
PN   WO2004110384-A2.
XX
PD   23-DEC-2004.
XX
PF   10-JUN-2004; 2004WO-US018672.
XX
PR   12-JUN-2003; 2003US-0477815P.
XX
PA   (VAXG-) VAXGEN INC.
XX
PI   Berman PW,  Jobes DV;
XX
DR   WPI; 2005-039971/04.
DR   P-PSDB; ADW69565.
XX
PT   Immunogenic composition useful for eliciting gp120-specific immune
PT   response, comprises isolated gp120 polypeptide or isolated polynucleotide
PT   encoding gp120 amino acid sequence.
XX
PS   Claim 44; SEQ ID NO 53; 130pp; English.
XX
CC   The invention relates to a novel immunogenic composition comprising an
CC   isolated polypeptide, or an isolated polynucleotide encoding, a gp120
CC   amino acid sequence, and a carrier, where the gp120 amino acid sequence
CC   comprises at least the V2, V3, and C4 domains of gp120. The invention
CC   further comprises: an isolated polypeptide comprising: a first gp120
CC   amino acid sequence having at least the V2, V3, and C4 domains of gp120,
CC   where the first gp120 sequence lacks one or more cysteine residues at one
CC   or more of the positions chosen from 54, 74, 119, 126, 157, 205, 218,
CC   228, 239, 247, 331, 378 and 385 and/or comprises one or more additional
CC   cysteine residues at a position other than the positions chosen from 54,
CC   74, 119, 126, 131, 157, 196, 205, 218, 228, 239, 247, 296, 331, 378, 385,
CC   418 and 445, provided that the one or more additional cysteine residues
CC   are not present in the V1 domain of gp120, as numbered from the N-
CC   terminal methionine of gp120 from the HXB-2 strain of HIV gp120, and the
CC   first gp120 sequence is not a subtype E gp120 sequence having one or more
CC   additional cysteines in the V4 domain; or a first gp120 amino acid
CC   sequence having at least 99% identity to each of the V1, V2, V3 and C4
CC   domains of a gp120 chosen from any one of 68 fully defined sequences of,
CC   e.g. U-101c1, U-178c13, U-013c6, U-016c1 and U-032 3 c4 amino acids, as
CC   given in the specification; an isolated polynucleotide encoding the
CC   isolated polypeptide; a vector comprising the isolated polynucleotide; a
CC   host cell comprising the vector; and a diagnostic method, comprising
CC   assaying a biological sample from a subject to determine whether the
CC   sample comprises a polypeptide comprising, or a polynucleotide encoding,
CC   a gp120 amino acid sequence that lacks one or more cysteine residues or
CC   comprises one or more additional cysteine residues at positions as
CC   mentioned above. The immunogenic composition has anti-HIV activity and
CC   may be used in the creation of a vaccine by eliciting a gp120-specific
CC   immune response. The immunogenic composition is useful for reducing the
CC   risk of, or preventing, infection of HIV-1 in individuals who are not
CC   already infected with HIV-1, for example: health professionals, police
CC   officers and fire fighters. This polynucleotide sequence represents the
CC   DNA encoding an HIV-1 gp120 protein of the invention.
XX
SQ   Sequence 1659 BP; 610 A; 270 C; 392 G; 387 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 14;  Length 1659;
  Best Local Similarity   69.0%;  Pred. No. 15;
  Matches   40;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy         71 CTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTTCAGC 128
              |||||| ||||  || |||   || |||| |||    ||| || | || | |||||||
Db        991 CTTTTTCTTGTATTGTTGTTGGGTCTTGTACAATTTAATTTTTACAGATTTGTTCAGC 934


RESULT 13
AEA79104
ID   AEA79104 standard; cDNA; 1760 BP.
XX
AC   AEA79104;
XX
DT   11-AUG-2005  (first entry)
```

```
XX
DE   Human apoptosis-associated cDNA SEQ ID 209.
XX
KW   apoptosis; cell death; cytostatic; neuroprotective; immunosuppressive;
KW   antirheumatic; antiarthritic; dermatological; antiinflammatory;
KW   hepatotropic; virucide; nootropic; anticonvulsant; antiparkinsonian;
KW   vasotropic; cerebroprotective; antialcoholic; gene therapy; tumour;
KW   autoimmune disease; degenerative disease; viral infection; leukaemia;
KW   carcinoma; sarcoma; multiple sclerosis; rheumatoid arthritis; diabetes;
KW   lupus; hepatitis; influenza viruses; Alzheimer's disease;
KW   Huntington's disease; Parkinson's diseases; reperfusion injury; stroke;
KW   alcoholic liver disease; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003058021-A2.
XX
PD   17-JUL-2003.
XX
PF   13-JAN-2003; 2003WO-EP000270.
XX
PR   11-JAN-2002; 2002DE-01000856.
XX
PA   (XANT-) XANTOS BIOMEDICINE AG.
XX
PI   Koenig-Hoffman K,  Kazinski M,  Schaefer R,  Kesper B;
XX
DR   WPI; 2003-542134/51.
XX
PT   New nucleic acids involved in apoptosis, useful for diagnosis and
PT   treatment of e.g. tumors and degenerative disease, also related proteins,
PT   antibodies and modulators.
XX
PS   Claim 1a; SEQ ID NO 209; 517pp; German.
XX
CC   This invention describes novel nucleic acid molecules that are associated
CC   with apoptosis and encode a polypeptide and are derived from a normalised
CC   gene library (embryonic or liver) or clone collections, and the extent of
CC   apoptosis measured by cell death detection assay or the CPRG assay
CC   (measuring loss of membrane integrity). The products of the invention
CC   have cytostatic, neuroprotective, immunosuppressive, antirheumatic,
CC   antiarthritic, dermatological, antiinflammatory, hepatotropic, virucide,
CC   nootropic, anticonvulsant, antiparkinsonian, vasotropic,
CC   cerebroprotective and antialcoholic activity and can be used for gene
CC   therapy. The polynucleotides also related vectors, hosts (or their
CC   extracts), encoded polypeptide (or their receptors) and/or agents that
CC   inhibit their activity (including antisense sequences) are used for
CC   treatment or prevention of tumours, autoimmune or degenerative diseases
CC   and viral infections, specifically leukaemia, carcinoma, sarcoma,
CC   multiple sclerosis, rheumatoid arthritis, diabetes, lupus, or infection
CC   with hepatitis or influenza viruses, Alzheimer's, Huntington's or
CC   Parkinson's diseases, reperfusion injury, stroke and alcoholic liver
CC   disease. Detection of the polynucleotides and derived polypeptides can
CC   also be used for diagnosis of these diseases. This sequence encodes an
CC   apoptosis-associated protein described in the disclosure of the
CC   invention.
XX
SQ   Sequence 1760 BP; 326 A; 513 C; 535 G; 386 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 11;  Length 1760;
  Best Local Similarity   59.8%;  Pred. No. 15;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
               | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db         762 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 821

Qy         108 ATTATTCCGGACTAGTTCAGCG 129
               || | | |  ||| | |   ||
Db         822 ATCAATGCCTACTTGGTGGACG 843


RESULT 14
```

```
ADR89493
ID   ADR89493 standard; DNA; 1800 BP.
XX
AC   ADR89493;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Apoptosis-inducing protein coding sequence, SEQ ID 17.
XX
KW   Cytostatic; Immunosuppressive; Neuroprotective; Vasotropic; Virucide;
KW   apoptosis; neurodegenerative disease; ischaemic disease; cancer;
KW   autoimmune disease; viral disease; murine; gene; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1170
FT                   /*tag=  a
FT                   /product= "Apoptosis-inducing protein"
XX
PN   WO2004078112-A2.
XX
PD   16-SEP-2004.
XX
PF   05-MAR-2004; 2004WO-JP002899.
XX
PR   07-MAR-2003; 2003JP-00061179.
PR   10-MAR-2003; 2003US-0452943P.
XX
PA   (ASAH-) ASAHI KASEI PHARMA CORP.
XX
PI   Muramatsu S,  Takeda M,  Matsuda A;
XX
DR   WPI; 2004-662343/64.
DR   P-PSDB; ADR89494.
XX
PT   New protein capable of inducing apoptosis, useful in screening for
PT   compounds that inhibit or induce apoptosis which may be used to treat
PT   neurodegenerative, ischemic, autoimmune and viral diseases, and cancer.
XX
PS   Claim 4; SEQ ID NO 17; 316pp; English.
XX
CC   The present invention relates to novel purified proteins (I) and their
CC   coding sequences (II) (ADR89477-ADR89550), which are capable of inducing
CC   apoptosis. The proteins (I) are useful as a target in screening for
CC   compounds that modulate apoptosis. Compounds that modulate the expression
CC   or activity of the protein may be useful in treating neurodegenerative
CC   diseases, ischaemic diseases, cancer, autoimmune diseases, or viral
CC   diseases. Measurement of the expression or activity of the protein may
CC   also be used to diagnose or disease or a susceptibility to a disease.
XX
SQ   Sequence 1800 BP; 323 A; 526 C; 556 G; 395 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 13;  Length 1800;
  Best Local Similarity   59.8%;  Pred. No. 15;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
               | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db         769 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 828

Qy         108 ATTATTCCGGACTAGTTCAGCG 129
               || | | |  ||| | |   ||
Db         829 ATCAATGCCTACTTGGTGGACG 850


RESULT 15
ABL01352
ID   ABL01352 standard; DNA; 1836 BP.
XX
AC   ABL01352;
XX
DT   15-MAR-2002  (first entry)
```

```
XX
DE   Murine apoptosis related DNA sequence #17.
XX
KW   Apoptosis; mouse; cancer; autoimmune disease; viral infection;
KW   Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW   reperfusion injury; stroke; liver damage; dilatory cardiomyopathy;
KW   transgenic animal; hepatotropic; antialcoholism; cytostatic;
KW   immunosuppressive; virucide; nootropic; neuroprotective; vasotropic;
KW   antiparkinsonian; cerebroprotective; ds.
XX
OS   Mus sp.
XX
PN   DE10126344-A1.
XX
PD   24-JAN-2002.
XX
PF   30-MAY-2001; 2001DE-01026344.
XX
PR   14-JUL-2000; 2000DE-01034303.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Grimm S,  Schoenfeld N,  Braziulis E,  Cramer U,  Gewies A,  Voss F;
PI   Mund T,  Albayrak T,  Gille H,  Klein M;
XX
DR   WPI; 2002-115563/16.
XX
PT   New apoptosis-associated nucleic acid sequences and polypeptides, useful
PT   for diagnosis, treatment and prevention of e.g. tumors and
PT   neurodegeneration.
XX
PS   Claim 1; Page 138-139; 227pp; German.
XX
CC   The present invention relates to nucleic acids from the mouse, where the
CC   nucleic acid is associated with apoptosis. The sequences can be used in
CC   the diagnosis, treatment and prevention of diseases associated with
CC   excessive or inadequate apoptosis, including tumours, autoimmune
CC   diseases, viral infections, degradative diseases (Alzheimer's,
CC   Parkinson's and Huntington's diseases), reperfusion injury, stroke and
CC   alcohol-induced injury to the liver, for identifying agents for treating
CC   these diseases, and to prepare transgenic animals in which expression of
CC   an apoptosis related sequence is altered. These are useful for genetic
CC   and/or pharmacological investigations of apoptosis and related diseases,
CC   including dilatory cardiomyopathy. The present sequence is one of the
CC   apoptosis related sequences of the invention
XX
SQ   Sequence 1836 BP; 341 A; 534 C; 562 G; 399 T; 0 U; 0 Other;

  Query Match             22.6%;  Score 29.2;  DB 6;  Length 1836;
  Best Local Similarity   59.8%;  Pred. No. 15;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        813 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 872

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        873 ATCAATGCCTACTTGGTGGACG 894


Search completed: March  7, 2006, 12:31:08
Job time : 307.617 secs
```

```
                    GenCore version 5.1.7
           Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 11:55:06 ; Search time 1586.98 Seconds
                                            (without alignments)
                                            3803.154 Million cell updates/sec

Title:          US-10-782-899-1_COPY_472_600
Perfect score:  129
Sequence:       1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters:      82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_est3:*
                4:  gb_htc:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_est7:*
                9:  gb_gss1:*
                10:  gb_gss2:*
                11:  gb_gss3:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                    %
Result              Query
  No.    Score      Match Length DB  ID                           Description
-------------------------------------------------------------------------------
 c  1     31.8      24.7     600  2  BG817831                     BG817831 ESSU0253
 c  2     31.4      24.3     786  8  CX773853                     CX773853 UI-EH-HG0
 c  3     31.2      24.2     748  8  CV949375                     CV949375 PVrpvb_13
    4       31      24.0     460  7  CF927863                     CF927863 laf58f02.
 c  5       31      24.0     932  7  CK279022                     CK279022 EST725100
 c  6     30.8      23.9     429  3  BM522110                     BM522110 ESSU0647
    7     30.8      23.9     524  1  AA517086                     AA517086 vh98f10.r
 c  8     30.8      23.9     938  7  CN808355                     CN808355 Blood_EST
 c  9     30.6      23.7     199  1  AJ282728                     AJ282728 4A3A-ABB-
   10     30.6      23.7     676  1  AV745983                     AV745983 AV745983
 c 11     30.6      23.7     790  10  CZ036908                     CZ036908 OM__Ba001
 c 12     30.2      23.4     244  7  CV609497                     CV609497 zcm31a01.
 c 13     30.2      23.4     494  3  BM522137                     BM522137 ESSU0674
 c 14     30.2      23.4     528  3  BM522043                     BM522043 ESSU0580
 c 15     30.2      23.4     539  3  BM522233                     BM522233 ESSU0771
 c 16     30.2      23.4     546  2  BG817800                     BG817800 ESSU0222
 c 17     30.2      23.4     572  3  BM522287                     BM522287 ESSU0825
 c 18     30.2      23.4     577  2  BG817916                     BG817916 ESSU0338
 c 19     30.2      23.4     599  2  BG817905                     BG817905 ESSU0327
 c 20     30.2      23.4     604  7  CV614668                     CV614668 zcn14a07.
 c 21     30.2      23.4     657  2  BG817854                     BG817854 ESSU0276
 c 22     30.2      23.4     682  2  BG817862                     BG817862 ESSU0284
 c 23     30.2      23.4     720  2  BG817630                     BG817630 ESSU0052
 c 24     30.2      23.4    1101  10  CNS017UC                     AL108510 Drosophil
 c 25     29.8      23.1     125  6  CA829999                     CA829999 3529_1_3_
 c 26     29.8      23.1     486  6  CB816450                     CB816450 3529_1_79
   27     29.8      23.1     655  7  CK561781                     CK561781 rswpb0_00
   28     29.8      23.1    1396  6  CD326834                     CD326834 AGENCOURT
 c 29     29.6      22.9     187  2  BF841751                     BF841751 RC2-HT107
 c 30     29.6      22.9     403  2  BG925693                     BG925693 HNC13-1-B
 c 31     29.6      22.9     819  7  CN168123                     CN168123 AGENCOURT
   32     29.6      22.9     903  11  CNS04GSE                     AL290039 Tetraodon
 c 33     29.6      22.9     937  6  CD107403                     CD107403 AGENCOURT
 c 34     29.4      22.8     137  3  BJ305960                     BJ305960 BJ305960
   35     29.4      22.8     364  4  AK210039                     AK210039 Mus muscu
 c 36     29.4      22.8     437  3  BM521973                     BM521973 ESSU0510
   37     29.4      22.8     485  8  H76502                       H76502 18207 Lambd
   38     29.4      22.8     770  10  CG997493                     CG997493 ZMMBBb053
   39     29.4      22.8    1353  2  BI114606                     BI114606 602861984
   40     29.2      22.6     231  4  AK178745                     AK178745 Mus muscu
   41     29.2      22.6     231  4  AK201694                     AK201694 Mus muscu
   42     29.2      22.6     231  4  AK208166                     AK208166 Mus muscu
   43     29.2      22.6     259  1  AI613789                     AI613789 vh98f10.y
   44     29.2      22.6     282  4  AK195241                     AK195241 Mus muscu
   45     29.2      22.6     282  4  AK199429                     AK199429 Mus muscu


                                    ALIGNMENTS


RESULT 1
BG817831/c
LOCUS       BG817831                 600 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  ESSU0253 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0334 5', mRNA sequence.
ACCESSION   BG817831
VERSION     BG817831.1  GI:14188811
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to pir|T17405 scavenger recept or cysteine-rich protein
            precursor - sea urchin (Strongyloce ntrotus purpuratus).
            Seq primer: T3 primer
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0334"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             24.7%;  Score 31.8;  DB 2;  Length 600;
```

```
  Best Local Similarity   57.6%;  Pred. No. 17;
  Matches   57;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy           2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
               |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db         423 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 364

Qy          62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
               | |  |||  | | || || |||  | ||   |||| ||
Db         363 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTTTTC 325


RESULT 2
CX773853/c
LOCUS       CX773853                 786 bp    mRNA    linear   EST 25-JAN-2005
DEFINITION  UI-EH-HG0-aaj-f-23-0-UI.s1 UI-EH-HG0 Emiliania huxleyi cDNA clone
            UI-EH-HG0-aaj-f-23-0-UI 3', mRNA sequence.
ACCESSION   CX773853
VERSION     CX773853.1  GI:58184206
KEYWORDS    EST.
SOURCE      Emiliania huxleyi
  ORGANISM  Emiliania huxleyi
            Eukaryota; Haptophyceae; Isochrysidales; Emiliania.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Provasoli-Guillard National Center for Culture
            of Marine Phytoplankton (CCMP)
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/dinoflagellate.html
             The following repetitive elements were found in this cDNA
            sequence: 148-189, >GC_rich#Low_complexity (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..786
                     /organism="Emiliania huxleyi"
                     /mol_type="mRNA"
                     /strain="CCMP 371"
                     /db_xref="taxon:2903"
                     /clone="UI-EH-HG0-aaj-f-23-0-UI"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-EH-HG0"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoR I; Site_2: Not I; The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site.  Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     GAAGGCTAGT. Tissue was obtained from the
                     Provasoli-Guillard National Center for Culture of Marine
                     Phytoplankton (CCMP).
                     TAG_TISSUE=Emiliania huxleyi coccolithophorid
                     TAG_LIB=UI-EH-HG0
                     TAG_SEQ=GAAGGCTAGT"
ORIGIN

  Query Match             24.3%;  Score 31.4;  DB 8;  Length 786;
  Best Local Similarity   56.2%;  Pred. No. 24;
  Matches   59;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy           4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
               || ||| | ||| ||   |  | ||||||   ||    || |||| | |   ||||  |
Db         339 CCCTCTTCACGTGGCTGGTGATCCGAGCAGCGCTCTACACCAAGAACGCCTACGAGCGGC 280

Qy          64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTA 108
                ||  | | |   | | ||| |  |  ||||| || | | ||| |
Db         279 GCGCCGGCATGACGCTCTGGAGGAGGCACGTGATGGCAATCCGCA 235


RESULT 3
CV949375/c
LOCUS       CV949375                 748 bp    mRNA    linear   EST 25-JAN-2005
DEFINITION  PVrpvb_13369 zoospores, purified Phytophthora infestans cDNA, mRNA
            sequence.
ACCESSION   CV949375
VERSION     CV949375.1  GI:58139131
KEYWORDS    EST.
SOURCE      Phytophthora infestans (potato late blight agent)
  ORGANISM  Phytophthora infestans
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   Randall,T., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C.,
            Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,E., Gaffney,T.,
            Law,M., Testa,A., Torto-Alalibo,A., Zhang,M., Zheng,L., Mueller,E.,
            Windass,J., Binder,A., Birch,P.R.J., Gisi,U., Govers,F., Gow,N.A.,
            Mauch,F., van West,P., Waugh,M.E., Yu,J., Boller,T., Kamoun,S.,
            Lam,S.T. and Judelson, H.S.
  TITLE     Large-scale gene discovery in the oomycete Phytophthora infestans
            reveals likely components of phytopathogenicity shared with true
            fungi
  JOURNAL   Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
   PUBMED   15782637
COMMENT     Contact: Judelson HS
            Department of Plant Pathology
            University of California
            Webber Hall, Riverside, CA 92521, USA
            Tel: 909 787 4199
            Fax: 909 787 4294
            Email: howard.judelson@ucr.edu.
FEATURES             Location/Qualifiers
     source          1..748
                     /organism="Phytophthora infestans"
                     /mol_type="mRNA"
                     /strain="88069"
                     /db_xref="taxon:4787"
                     /sex="A1"
                     /clone_lib="zoospores, purified"
                     /note="Vector: pSPORT1"
ORIGIN

  Query Match             24.2%;  Score 31.2;  DB 8;  Length 748;
  Best Local Similarity   53.2%;  Pred. No. 27;
  Matches   66;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy           2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
               || |||    || | ||||  ||        ||   ||||   |      |||  ||| |
Db         145 GAACTTGACGTCCTCGCGACGACCGTCGCGCTTGACGACGTACATCGTTGCGATTGAGCT 86

Qy          62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTA 121
               ||||||   | |  | | | ||||    |||| || |  | |||     |||||||||
Db          85 GGTCGGTAACGTGCTCTGGAGGTGGACGGACGCGTGGGTCGACCCGGGAATTCCGGACCG 26

Qy         122 GTTC 125
               || |
Db          25 GTAC 22
```

```
RESULT 4
CF927863
LOCUS       CF927863                 460 bp    mRNA    linear   EST 12-NOV-2003
DEFINITION  laf58f02.y1 SiEP Mus musculus cDNA 5' similar to TR:Q9UJZ5 Q9UJZ5
            PRESENILIN-ASSOCIATED PROTEIN. ;, mRNA sequence.
ACCESSION   CF927863
VERSION     CF927863.1  GI:38274486
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Tidwell,R., Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J.,
            Wylie,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennet,J.,
            Ronko,I., Tsagareishvili,R., Belaygorod,L., Grow,A., Maguire,L.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Stem cell EST Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Jeff Gordon and Mike Lovett
            WashU, Human Genetics Division
            Washington University School of Medicine
            Library materials provided by: Stappenbeck Th.S. and Gordon J.I.
            Library constructed by: Korshunova Y. and Lovett M. DNA sequencing
            by: Washington University Genome Sequencing Center For information
            on obtaining a clone please contact: Rose Tidwell
            (rtidwell@genetics.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="small intestinal epithelial progenitors"
                     /lab_host="DH5alpha Ultra Max cells (Invitrogen)"
                     /clone_lib="SiEP"
                     /note="Vector: pAMP1; The library was synthesized with
                     modified SMART primers with dUTP at the end. After
                     treatment with UDG the cDNA was cloned in pAMP1 vector by
                     annealing."
ORIGIN

  Query Match             24.0%;  Score 31;  DB 7;  Length 460;
  Best Local Similarity   59.8%;  Pred. No. 29;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         43 CAAAGACCCCGACCGAGATGGTCGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCA 102
              | |  | |||   | | | || ||  ||  |||| |||||| ||||| || || || ||
Db          7 CTATTAACCCTCACTAAAGGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCC 66

Qy        103 ACCGTATTATTCCGGACTAGTTCAGCG 129
              ||   || | | |  ||| | |   ||
Db         67 ACTTCATCAATGCCTACTTGGTGGACG 93


RESULT 5
CK279022/c
LOCUS       CK279022                 932 bp    mRNA    linear   EST 03-AUG-2004
DEFINITION  EST725100 potato abiotic stress cDNA library Solanum tuberosum cDNA
            clone POAE963 5' end, mRNA sequence.
ACCESSION   CK279022
VERSION     CK279022.1  GI:39836000
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
```

```
  TITLE     Generation of ESTs from abiotic stressed potato tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from the University of Arizona Genomics
            Institute via http://genome.arizona.edu/orders/ .
            Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES             Location/Qualifiers
     source          1..932
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="POAE963"
                     /tissue_type="abiotic stress treated leaf and root tissue"
                     /lab_host="DH10B-TonA"
                     /clone_lib="potato abiotic stress cDNA library"
                     /note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI;
                     supplier: Solanum tuberosum var. Kennebec plants were
                     grown from cuttings on a 16hr light/8 hr dark cycle at 25
                     C for 3-4 weeks. Abiotic stress conditions were applied to
                     four separate sets of plants. Set 1 involved saturation of
                     the soil with 150 mM NaCl and tissues were harvested at
                     following application of the salt stress (leaves: 2hr,
                     6hr, 12hr, 1d, 2d, and 4d; roots:2hr, 6hr, 12hr, and 2d).
                     Set 2 were grown under the standard conditions and then
                     were water stressed by withdrawal of further watering
                     applications. Drought stressed plants were harvested after
                     cessation of watering (leaves: 3d, 5d, and 7d; roots:3d
                     and 5d). Set 3 were grown under the standard conditions
                     and then were cold stressed by placement at 4 C. Cold
                     stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     2d. Set 4 were grown under the standard conditions and
                     then were heat stressed by placement at 35 C. Heat
                     stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
                     2d and 4d and heat-stressed roots were harvested at 6 hr,
                     12 hr, 1 d, and 4d. RNA was isolated from all tissues and
                     equal RNA from each tissue and stress was pooled to
                     construct the cDNA library. RNA sample."
ORIGIN

  Query Match             24.0%;  Score 31;  DB 7;  Length 932;
  Best Local Similarity   57.9%;  Pred. No. 33;
  Matches   55;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          5 CTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGT 64
              |||||| |  ||  || ||    |   || | ||| || |||||    ||  | |||| |
Db        844 CTTCTGTTAATATTGAATAGAAAGCTGATGATTGAGGAGAAAGAGAGTGATGGTGATGAT 785

Qy         65 CGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGT 99
                 | | || ||||||| ||  |      ||||||
Db        784 GTTGTTGTTGTTGTTGTTGTAATTGATGTTGTTGT 750


RESULT 6
BM522110/c
LOCUS       BM522110                 429 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0647 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0777 5', mRNA sequence.
ACCESSION   BM522110
VERSION     BM522110.1  GI:18706032
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
```

```
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7495754|pir|T19130 hypothetical protein C09F9.2
            [Caenorhabditis elegans]
            Seq primer: T3 primer
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0777"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match            23.9%;  Score 30.8;  DB 3;  Length 429;
  Best Local Similarity  57.1%;  Pred. No. 33;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        406 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 347

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGT 99
              | |  |||  | | || || |||  | ||   |||| |
Db        346 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTTTT 309


RESULT 7
AA517086
LOCUS       AA517086                 524 bp    mRNA    linear   EST 14-JUL-1997
DEFINITION  vh98f10.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:902347 5', mRNA sequence.
ACCESSION   AA517086
VERSION     AA517086.1  GI:2256675
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:523011
            Seq primer: -28m13 rev2 ET from Amersham
```

```
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:902347"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.   "
ORIGIN

  Query Match            23.9%;  Score 30.8;  DB 1;  Length 524;
  Best Local Similarity  61.0%;  Pred. No. 35;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||| |||| |||||| ||||| || || || || ||
Db         36 ATCCCTCACCTCCTGGGCGATGTCGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 95

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db         96 ATCAATGCCTACTTGCTGCACG 117


RESULT 8
CN808355/c
LOCUS       CN808355                 938 bp    mRNA    linear   EST 27-MAY-2004
DEFINITION  Blood_EST0439 Metarhizium anisopliae ARSEF 2575 from insect blood
            Metarhizium anisopliae cDNA clone B819 5', mRNA sequence.
ACCESSION   CN808355
VERSION     CN808355.1  GI:47729828
KEYWORDS    EST.
SOURCE      Metarhizium anisopliae
  ORGANISM  Metarhizium anisopliae
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
            Clavicipitaceae; Metarhizium.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   Wang,C.S., Hu,G. and St. Leger,R.J.
  TITLE     Gene expression profiling of Metarhizium anisopliae grown under
            different conditions: mechanisms of fungal opportunism
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Wang CS
            Department of Entomology
            University of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Email: cwang4@umd.edu
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Metarhizium anisopliae"
                     /mol_type="mRNA"
                     /db_xref="taxon:5530"
                     /clone="B819"
                     /clone_lib="Metarhizium anisopliae ARSEF 2575 from insect
                     blood"
                     /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was
                     grown in insect haemolymph for 24 hours. A cDNA library
                     was constructed in the vector pCMV.SPORT6.1"
ORIGIN
```

```
  Query Match             23.9%;  Score 30.8;  DB 7;  Length 938;
  Best Local Similarity   63.5%;  Pred. No. 38;
  Matches   47;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          7 TCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCG 66
              ||||  ||  | |  | ||||  || || |||||||||||| |  || |||   |||
Db        330 TCTGGGCGGCGGGGGTGCTTCCGGCGTTCCTGACGACAAAGTCGACGGCCGTCTTGGAGT 271

Qy         67 GGGTCTTTTTGTTG 80
               | ||||   ||||
Db        270 AGATCTTGGGGTTG 257


RESULT 9
AJ282728/c
LOCUS       AJ282728                 199 bp    mRNA    linear   EST 30-JUN-2000
DEFINITION  4A3A-ABB-D-05-R Anopheles gambiae immune competent 4A3A Anopheles
            gambiae cDNA clone 4A3A-ABB-D-05, mRNA sequence.
ACCESSION   AJ282728
VERSION     AJ282728.1  GI:6930607
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae pilot gene discovery project: identification of
            mosquito innate immunity genes from expressed sequence tags
            generated from immune-competent cell lines
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
   PUBMED   10841561
COMMENT     Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES             Location/Qualifiers
     source          1. .199
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="4A r/r"
                     /db_xref="taxon:7165"
                     /clone="4A3A-ABB-D-05"
                     /cell_line="immune competent 4A3A"
                     /lab_host="E. coli DH10B"
                     /clone_lib="Anopheles gambiae immune competent 4A3A"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from
                     forward priming site which reads from the 3' end of the
                     cDNA. The 4A3A is a directionally cloned and normalized
                     cDNA library that was constructed from the 4A3A cell line
                     oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
                     (1996) : Normalization and Subtraction: Two approaches To
                     Facilitate Gene Discovery, Genome Research 6, 791-806."
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 1;  Length 199;
  Best Local Similarity   55.0%;  Pred. No. 34;
  Matches   60;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         12 TCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTC 71
              | | ||| |||     | ||| |||| |||  || |  ||    ||  |||  | |||
Db        184 TGGGAGCAATTTAACGGCGCACTACTAACGGTAACGGACCAACACGTAATGTCCCGGGGT 125

Qy         72 TTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
               |||| || || | |  |     || ||  || ||  || | |  ||||
Db        124 CTTTTTTTTTGTTTCATTTGGCGGTAGTAAAATCGGCTTCTACTAGACT 76
```

```
RESULT 10
AV745983
LOCUS       AV745983                 676 bp    mRNA    linear   EST 22-NOV-2004
DEFINITION  AV745983 NPA Homo sapiens cDNA clone NPAAZG08 5', mRNA sequence.
ACCESSION   AV745983
VERSION     AV745983.2  GI:55951951
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Song,H., Peng,Y., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
            Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S.,
            Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Ye,M.,
            Zhang,Q., Han,Z., Chen,Z., Hu,R. and Chen,J.
  TITLE     Homo sapiens NPA library cDNA clones
  JOURNAL   Unpublished (2000)
COMMENT     On Oct 18, 2000 this sequence version replaced gi:10865430.
            Contact: Qinghua Zhang
            Shanghai Institute of Endocrinology, Rui-Jin Hospital
            197 Rui-Jin II Road, Shanghai 200025, P. R. China
            Tel: 86-21-64370045(ex.663332)
            Fax: 86-21-64743206
            Email: mbshi@ms.stn.sh.cn
            This clone is available at Shanghai Hematology Institute in
            Shanghai.
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi- Tech Park, Pudong.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NPAAZG08"
                     /tissue_type="pituitary"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="NPA"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 1;  Length 676;
  Best Local Similarity   53.8%;  Pred. No. 42;
  Matches   63;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||  || || |         |  |   | |  | || ||||  || |    | ||||
Db        435 GACCGCCTCCTGGATCTCCCCGCCGCAGCCTTCTCAGAAGACATCGAGCGGTCCTGAGAG 494

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGA 118
                || |||    |||||    ||||||| || ||  ||| |||| ||  ||   |||
Db        495 CCTCTGGGCACGTTTGTCTGTGTGCTGTAACCTGAAGTCAAACCTTAAGATAATGGA 551


RESULT 11
CZ036908/c
LOCUS       CZ036908                 790 bp    DNA     linear   GSS 10-JAN-2005
DEFINITION  OM__Ba0016I22.f OM__Ba Oryza minuta genomic clone OM__Ba0016I22 5',
            genomic survey sequence.
ACCESSION   CZ036908
VERSION     CZ036908.1  GI:57380375
KEYWORDS    GSS.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
```

```
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            Plate: 0016  row: I  column: 22
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..790
                     /organism="Oryza minuta"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:63629"
                     /clone="OM__Ba0016I22"
                     /tissue_type="leaves"
                     /lab_host="DH10B"
                     /clone_lib="OM__Ba"
                     /note="Vector: pCUGIBAC1; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 10;  Length 790;
  Best Local Similarity   60.0%;  Pred. No. 43;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         42 ACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCC 101
              || | ||   | | ||| |  |||||    ||||| |||   ||| || |||||  || |
Db        510 ACGACGAAGACCATCGAAAATGTCGGCTAGTTTTTTTTGGCTTGCGGTCACGTGCGGTAC 451

Qy        102 AACCGTATTATTCCGGACTAGTTCA 126
              ||| |||| | |  |    | || |
Db        450 AACTGTATAAATAAGTGGAATTTAA 426


RESULT 12
CV609497/c
LOCUS       CV609497                 244 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  zcm31a01.y2 F_HC7CDNA Ajellomyces capsulatus cDNA 5', mRNA
            sequence.
ACCESSION   CV609497
VERSION     CV609497.1  GI:54597079
KEYWORDS    EST.
SOURCE      Ajellomyces capsulatus
  ORGANISM  Ajellomyces capsulatus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; Ajellomycetaceae; Ajellomyces.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   Magrini,V., Sil,A., Goldman,W. and Mardis,R.
  TITLE     Genomic Resources for Histoplasma Sequencing Mardis/WashU
            Histoplasma capsulatum Sequencing Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Vincent Magrini
            Mardis/Goldman/Sil Genomic Resources for Histoplasma Sequencing
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: vmagrini@watson.wustl.edu
            Contact Vincent Magrini (vmagrini@watson.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: H7cDNA-F2.
FEATURES             Location/Qualifiers
     source          1..244
                     /organism="Ajellomyces capsulatus"
                     /mol_type="mRNA"
                     /strain="G217B"
                     /db_xref="taxon:5037"
                     /clone_lib="F_HC7CDNA"
                     /note="Vector: pAS209; Site_1: NotI; Site_2: SalI;
                     Histoplasma G217B cDNA library."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 7;  Length 244;
  Best Local Similarity   62.7%;  Pred. No. 48;
  Matches   47;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy         13 CGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCT 72
              ||||| || ||| ||     ||| ||| | || ||||  |  ||      ||   |||||
Db        233 CGTAGAGAGTACGTCCTTGTTTATTGAGGGCATAGACAACATCCAGCGATGTGTCGGTCT 174

Qy         73 TTTTGTTGTGGTGCT 87
              || | |||| |||||
Db        173 TTCTCTTGTCGTGCT 159


RESULT 13
BM522137/c
LOCUS       BM522137                 494 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0674 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0807 5', mRNA sequence.
ACCESSION   BM522137
VERSION     BM522137.1  GI:18706191
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7506864|pir|T24294 hypothetical protein T01D3.3b -
            [Caenorhabditis elegans]
            Seq primer: T3 primer
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1..494
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0807"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 3;  Length 494;
  Best Local Similarity   56.6%;  Pred. No. 54;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        135 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 76

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db         75 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 37
```

```
RESULT 14
BM522043/c
LOCUS       BM522043                 528 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0580 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0706 5', mRNA sequence.
ACCESSION   BM522043
VERSION     BM522043.1  GI:18705601
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7512096|pir|T17405 scavenger receptor cysteine-rich
            protein precursor - sea urchin [Strongylocentrotus purpuratus]
            Seq primer: T3 primer
            High quality sequence stop: 528.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0706"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 3;  Length 528;
  Best Local Similarity   56.6%;  Pred. No. 55;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        387 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 328

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        327 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 289


RESULT 15
BM522233/c
LOCUS       BM522233                 539 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0771 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
            SAS0908 5', mRNA sequence.
ACCESSION   BM522233
VERSION     BM522233.1  GI:18706995
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Similar to gi|7512096|pir|T17405 scavenger receptor cysteine-rich
            protein precursor - sea urchin [Strongylocentrotus purpuratus]
            Seq primer: T3 primer
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0908"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

  Query Match             23.4%;  Score 30.2;  DB 3;  Length 539;
  Best Local Similarity   56.6%;  Pred. No. 55;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              |||| |  |  |||    ||    ||| |||| |  ||| |||| || ||| |     ||
Db        402 GACCATGAGGACGTGCACATGTTCTCGGGCATGAACGACCACAAGGAGCCCAATGAGAAT 343

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
              | |  |||  | | || || |||  | ||   |||  ||
Db        342 GTTTAGGGCATCTGTGGTGAGGTCTTTTGCAATGTCTTC 304


Search completed: March  7, 2006, 13:19:43
Job time : 1595.98 secs
```

```
                         GenCore version 5.1.7
                Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:        March  7, 2006, 11:56:56 ; Search time 90.4206 Seconds
                                          (without alignments)
                                          2535.988 Million cell updates/sec

Title:         US-10-782-899-1_COPY_472_600
Perfect score: 129
Sequence:      1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters:      2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
               1:  /cgn2_6/ptodata/1/ina/1_COMB.seq:*
               2:  /cgn2_6/ptodata/1/ina/5_COMB.seq:*
               3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
               4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
               5:  /cgn2_6/ptodata/1/ina/H_COMB.seq:*
               6:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
               7:  /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
               8:  /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
               9:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result             Query
  No.    Score     Match Length DB  ID                      Description
  -----------------------------------------------------------------------
     1    29.4     22.8    870  3  US-09-583-110-1324      Sequence 1324, Ap
     2    29.4     22.8    885  3  US-09-107-433-306       Sequence 306, App
     3    29.4     22.8 194915  3  US-09-949-016-15584     Sequence 15584, A
     4      29     22.5    601  3  US-09-949-016-68059     Sequence 68059, A
     5      29     22.5 264665  3  US-09-949-016-13747     Sequence 13747, A
 c   6    28.6     22.2   1146  3  US-09-252-991A-2602     Sequence 2602, Ap
     7    28.6     22.2   2184  3  US-09-252-991A-2241     Sequence 2241, Ap
     8    28.6     22.2   2868  3  US-09-252-991A-2311     Sequence 2311, Ap
     9    28.4     22.0    936  2  US-08-602-359A-31       Sequence 31, Appl
 c  10    27.4     21.2   1160  3  US-09-270-767-13705     Sequence 13705, A
    11    27.4     21.2   1951  3  US-10-144-929-56        Sequence 56, Appl
 c  12    26.6     20.6    601  3  US-09-949-016-125432    Sequence 125432,
    13    26.6     20.6 139552  3  US-09-949-016-15300     Sequence 15300, A
 c  14    26.4     20.5  14707  3  US-09-312-762A-3        Sequence 3, Appli
 c  15    26.2     20.3    601  3  US-09-949-016-143314    Sequence 143314,
 c  16    26.2     20.3    601  3  US-09-949-016-143485    Sequence 143485,
 c  17    26.2     20.3    746  3  US-09-380-419C-1        Sequence 1, Appli
 c  18    26.2     20.3    907  3  US-08-672-850-7         Sequence 7, Appli
 c  19    26.2     20.3    907  3  US-08-672-850-11        Sequence 11, Appl
 c  20    26.2     20.3    907  3  US-09-565-177A-7        Sequence 7, Appli
 c  21    26.2     20.3    907  3  US-09-565-177A-11       Sequence 11, Appl
 c  22    26.2     20.3   1001  3  US-08-672-850-10        Sequence 10, Appl
 c  23    26.2     20.3   1001  3  US-09-565-177A-10       Sequence 10, Appl
    24    26.2     20.3  85912  3  US-09-949-016-12362     Sequence 12362, A
```

```
    25    26.2     20.3  85913  3  US-09-949-016-16109     Sequence 16109, A
    26    26.2     20.3 120727  3  US-09-949-016-15787     Sequence 15787, A
    27    26.2     20.3 120727  3  US-09-949-016-15788     Sequence 15788, A
 c  28      26     20.2   1074  3  US-09-934-901-7         Sequence 7, Appli
 c  29      26     20.2   1074  3  US-09-934-868-17        Sequence 17, Appl
 c  30      26     20.2   1074  3  US-10-321-210-7         Sequence 7, Appli
 c  31      26     20.2   1074  3  US-10-320-874-7         Sequence 7, Appli
 c  32      26     20.2   1074  3  US-10-320-924-7         Sequence 7, Appli
 c  33      26     20.2   1074  3  US-10-701-200-17        Sequence 17, Appl
 c  34      26     20.2  43280  2  US-08-804-227C-1        Sequence 1, Appli
    35      26     20.2 336024  3  US-09-949-016-12373     Sequence 12373, A
 c  36    25.8     20.0   1158  3  US-09-198-092-1         Sequence 1, Appli
 c  37    25.8     20.0   2262  2  US-08-674-887A-5        Sequence 5, Appli
 c  38    25.8     20.0   2262  3  US-08-951-844-5         Sequence 5, Appli
 c  39    25.8     20.0   2262  3  US-09-412-347-5         Sequence 5, Appli
    40    25.6     19.8    271  3  US-09-313-294A-1215     Sequence 1215, Ap
 c  41    25.6     19.8   6413  3  US-08-956-991-10        Sequence 10, Appl
 c  42    25.6     19.8   6604  3  US-08-956-991-1         Sequence 1, Appli
    43    25.6     19.8  46725  3  US-09-949-016-15680     Sequence 15680, A
    44    25.6     19.8 183112  3  US-09-949-016-14184     Sequence 14184, A
 c  45    25.6     19.8 481115  3  US-09-949-016-11940     Sequence 11940, A


                                 ALIGNMENTS


RESULT 1
US-09-583-110-1324
; Sequence 1324, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 1324
;   LENGTH: 870
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-1324

  Query Match             22.8%;  Score 29.4;  DB 3;  Length 870;
  Best Local Similarity   58.6%;  Pred. No. 1.2;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db        589 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 648

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        649 CGTGGTTTTGCCTGGTTGGATACGGGA 675


RESULT 2
US-09-107-433-306
; Sequence 306, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGN
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
```

```
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 306:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 885 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...885
;        SEQUENCE DESCRIPTION: SEQ ID NO: 306:
US-09-107-433-306

  Query Match             22.8%;  Score 29.4;  DB 3;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 1.2;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
               ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
Db         604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy          92 CGTGTTGTCCAACCGTATTATTCCGGA 118
               |||| | |      ||   || | |||
Db         664 CGTGGTTTTGCCTGGTTGGATACGGGA 690


RESULT 3
US-09-949-016-15584
; Sequence 15584, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
```

```
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15584
;   LENGTH: 194915
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(194915)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15584

  Query Match             22.8%;  Score 29.4;  DB 3;  Length 194915;
  Best Local Similarity   60.8%;  Pred. No. 11;
  Matches   48;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          42 ACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCC 101
               | |||  | ||    ||||     ||   || |||||||||||||| ||  || ||| |
Db      115691 AGAAACTCACCACTTGAGACAAAAGTGAACTCTTTGTTGTGGTGCTATGTGGTCTTGGCA 115750

Qy         102 AACCGTATTATTCCGGACT 120
                |   |  | |||||| ||
Db      115751 GAGGTTTGTTTTCCGGTCT 115769


RESULT 4
US-09-949-016-68059
; Sequence 68059, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68059
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-68059

  Query Match             22.5%;  Score 29;  DB 3;  Length 601;
  Best Local Similarity   58.8%;  Pred. No. 1.4;
  Matches   50;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          23 ACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTG 82
               | | | || ||||| | ||  |   |||  | |||    |  | || ||||| | ||||
Db          30 AGTGCTAGGATTACAGGCGTGAGCCACCGTGCCCGGCCAGACCTGGCTCTTTCTTTTGTT 89

Qy          83 GTGCTGTGACGTGTTGTCCAACCGT 107
                |  || ||||   |||||  | ||
Db          90 TTTTTGAGACGGAGTGTCCCTCTGT 114


RESULT 5
US-09-949-016-13747
; Sequence 13747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
```

```
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13747
;   LENGTH: 264665
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13747

  Query Match             22.5%;  Score 29;  DB 3;  Length 264665;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   50;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         23 ACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTG 82
              | | | || ||||| | ||  |   |||  | |||    |  | || ||||| | ||||
Db      52981 AGTGCTAGGATTACAGGCGTGAGCCACCGTGCCCGGCCAGACCTGGCTCTTTCTTTTGTT 53040

Qy         83 GTGCTGTGACGTGTTGTCCAACCGT 107
               |  || ||||   |||||  | ||
Db      53041 TTTTTGAGACGGAGTGTCCCTCTGT 53065


RESULT 6
US-09-252-991A-2602/c
; Sequence 2602, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2602
;   LENGTH: 1146
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2602

  Query Match             22.2%;  Score 28.6;  DB 3;  Length 1146;
  Best Local Similarity   57.1%;  Pred. No. 2.6;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db        906 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 847

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     |  |    | || || |||
Db        846 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 816


RESULT 7
US-09-252-991A-2241
; Sequence 2241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
```

```
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2241
;   LENGTH: 2184
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2241

  Query Match             22.2%;  Score 28.6;  DB 3;  Length 2184;
  Best Local Similarity   57.1%;  Pred. No. 3.4;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db       1538 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1597

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     |  |    | || || |||
Db       1598 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1628


RESULT 8
US-09-252-991A-2311
; Sequence 2311, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 2311
;   LENGTH: 2868
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2311

  Query Match             22.2%;  Score 28.6;  DB 3;  Length 2868;
  Best Local Similarity   57.1%;  Pred. No. 3.8;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy         39 ACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTG 98
              ||||  ||||||| | ||||  | |  | | ||||  || |||  |||   | | ||  |
Db       1177 ACGATCAAGACCCTGGCCGAACTCGGGGTGATCTTCCTGATGTTCTGCCTCGGCCTGGAG 1236

Qy         99 TCCAACCGTATTATTCCGGACTAGTTCAGCG 129
              | || ||     |  |    | || || |||
Db       1237 TTCAGCCTGCGCAAGCTCTTCCAGGTCGGCG 1267


RESULT 9
US-08-602-359A-31
; Sequence 31, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERTSON, Daniel E.
;    APPLICANT:  MURPHY, Dennis
;    APPLICANT:  REID, John
;    APPLICANT:  MAFFIA, Anthony
;    APPLICANT:  LINK, Steven
;    APPLICANT:  SWANSON, Ronald V.
;    APPLICANT:  WARREN, Patrick V.
```

```
;    APPLICANT:  KOSMOTKA, Anna
;    TITLE OF INVENTION:  ESTERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  GENOMIC DNA
US-08-602-359A-31

  Query Match             22.0%;  Score 28.4;  DB 2;  Length 936;
  Best Local Similarity   56.4%;  Pred. No. 2.8;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||||  |  | || |   | |   || ||| || | |||  | |||||     |||||
Db        718 GACCTTGAGAACCTACCTCCTGCGCTGATCATAACCGCCGAATACGACCCGCTGAGAGAT 777

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
              |   | |   |||| |    | ||||| || | |
Db        778 GAAGGAGAAGTTTTCGGGCAGATGCTGAGAAGAG 811


RESULT 10
US-09-270-767-13705/c
; Sequence 13705, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13705
;   LENGTH: 1160
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-13705

  Query Match             21.2%;  Score 27.4;  DB 3;  Length 1160;
  Best Local Similarity   54.5%;  Pred. No. 7.2;
  Matches   55;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;
```

```
Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || | | |  || |    || ||| | |     ||   | ||| ||| ||| |  ||||
Db        656 CCATTTTCGAGTCGTCGCTAATTCCATCCACCTTGTGTATAAACACCTCGAACTTGATGC 597

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAAC 104
               |  | ||  ||||| |   ||| |    ||||| |  |||
Db        596 GCTTGTTCACTTTGTAGGCCTGCAGCACTGTGTTCTTGAAC 556


RESULT 11
US-10-144-929-56
; Sequence 56, Application US/10144929
; Patent No. 6881823
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 70 Human Secreted Proteins
;  FILE REFERENCE: PZ014P1
;  CURRENT APPLICATION NUMBER: US/10/144,929
;  CURRENT FILING DATE:  2002-05-15
;  PRIOR APPLICATION NUMBER: US/09/251,329
;  PRIOR FILING DATE: 1999-02-17
;  PRIOR APPLICATION NUMBER: PCT/US98/17044
;  PRIOR FILING DATE: 1998-08-18
;  NUMBER OF SEQ ID NOS: 257
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
;   LENGTH: 1951
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (28)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1636)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1947)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1951)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-144-929-56

  Query Match             21.2%;  Score 27.4;  DB 3;  Length 1951;
  Best Local Similarity   54.5%;  Pred. No. 8.9;
  Matches   55;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         24 CTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGG 83
              || |  | |||  |      |   | |||   |    ||| ||  ||  |||| ||||||
Db        718 CTGCTGGGATTCTTCGTTGGATTAATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGG 777

Qy         84 TGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACTAGTT 124
               ||||| || || || || ||   || | | |  ||  | |
Db        778 GGCTGTAACCTGCTGGCCCACTTCATCAATGCCTACCTGGT 818


RESULT 12
US-09-949-016-125432/c
; Sequence 125432, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
```

```
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 125432
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-125432

  Query Match             20.6%;  Score 26.6;  DB 3;  Length 601;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches   38;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          44 AAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
               |||| | | ||  ||||||     |||||| |   | ||| ||||| | || |||||
Db         407 AAAGCCTCTGATTGAGATGAGGAAGGTCTTCTACCTCTGGAGCTGTAAGGTCTTGTC 351


RESULT 13
US-09-949-016-15300
; Sequence 15300, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15300
;   LENGTH: 139552
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15300

  Query Match             20.6%;  Score 26.6;  DB 3;  Length 139552;
  Best Local Similarity   66.7%;  Pred. No. 99;
  Matches   38;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          44 AAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTC 100
               |||| | | ||  ||||||     |||||| |   | ||| ||||| | || |||||
Db       80719 AAAGCCTCTGATTGAGATGAGGAAGGTCTTCTACCTCTGGAGCTGTAAGGTCTTGTC 80775


RESULT 14
US-09-312-762A-3/c
; Sequence 3, Application US/09312762A
; Patent No. 6552177
;  GENERAL INFORMATION:
;    APPLICANT:  MIA HOROWITZ ET AL.
;    TITLE OF INVENTION:  EH DOMAIN CONTAINING GENES AND PROTEINS
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mark M. Friedman c/o Anthony Castorina
;      STREET:  2001 Jefferson Davis Highway, Suite 207
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  1.44 megabyte, 3.5" microdisk
;      COMPUTER:  Twinhead* Slimnote-890TX
```

```
;      OPERATING SYSTEM:  MS DOS version 6.2,
;      OPERATING SYSTEM:  Windows version 3.11
;      SOFTWARE:  Word for Windows version 2.0 converted to
;      SOFTWARE:  an ASCI file
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/312,762A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/026,898
;      FILING DATE:  20 FEB 1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Friedmam, Mark M.
;      REGISTRATION NUMBER:  33,883
;      REFERENCE/DOCKET NUMBER:  916/10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  972-3-5625553
;      TELEFAX:  972-3-5625554
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14707
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-312-762A-3

  Query Match             20.5%;  Score 26.4;  DB 3;  Length 14707;
  Best Local Similarity   54.0%;  Pred. No. 47;
  Matches   54;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy           7 TCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCG 66
               |||  || | | ||||  |  | || ||  ||||  |  ||| | || | ||||||  |
Db        5137 TCTTGTCCTCGTGATTTTTGAGGGCCTTGATGACTTCTGAGAACTCGTCTGAGATGTCCA 5078

Qy          67 GGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCG 106
               |  | |    || |     |   ||  ||  |||||  ||
Db        5077 GCTTGTGGGCGTCGAACAACAAAGAGATGCGGTCCACACG 5038


RESULT 15
US-09-949-016-143314/c
; Sequence 143314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 143314
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-143314

  Query Match             20.3%;  Score 26.2;  DB 3;  Length 601;
  Best Local Similarity   58.2%;  Pred. No. 15;
  Matches   46;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          29 AGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTG 88
               ||| |  |||||  | | | || | || |   |||| |||  |  | |     | | |||
Db         291 AGCTTCTCTGACACCCACGCCCACCACAGTCTTGGTTGGGTGCCCTGTCCCCAGATCCTG 232
```

```
Qy         89 TGACGTGTTGTCCAACCGT 107
              || | | |||| || | ||
Db        231 TGGCTTTTTGTTCACCTGT 213

Search completed: March  7, 2006, 14:11:48
Job time : 93.4206 secs
```

```
                     GenCore version 5.1.7
          Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:          March  7, 2006, 12:18:50 ; Search time 1192.75 Seconds
                                            (without alignments)
                                            237.125 Million cell updates/sec

Title:           US-10-782-899-1_COPY_472_600
Perfect score:   129
Sequence:        1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        7218535 seqs, 1096242582 residues

Total number of hits satisfying chosen parameters:      14437070

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_New:*
                 1:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
                 7:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
                 9:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
                 11:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
                 12:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
                 13:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                 %
Result           Query
  No.   Score    Match Length DB  ID                            Description
 ------------------------------------------------------------------------------
    1      29     22.5 185750  9  US-11-114-798-56              Sequence 56, Appl
    2      29     22.5 1080000  8  US-10-928-446A-1              Sequence 1, Appli
    3      29     22.5 1080000  8  US-10-928-446A-181            Sequence 181, App
    4      29     22.5 1080000  8  US-10-928-446A-183            Sequence 183, App
    5      29     22.5 1080000  8  US-10-928-446A-185            Sequence 185, App
    6      29     22.5 1080000  8  US-10-928-446A-187            Sequence 187, App
    7      29     22.5 1080000  8  US-10-928-446A-189            Sequence 189, App
    8      29     22.5 1080000  8  US-10-928-446A-191            Sequence 191, App
    9      29     22.5 1080000  8  US-10-928-446A-193            Sequence 193, App
   10      29     22.5 1080000  8  US-10-928-446A-195            Sequence 195, App
   11      29     22.5 1080000  8  US-10-928-446A-197            Sequence 197, App
   12      29     22.5 1080000  8  US-10-928-446A-199            Sequence 199, App
   13      29     22.5 1080000  8  US-10-928-446A-201            Sequence 201, App
c  14    28.8     22.3  79134 12  US-11-124-368A-2924           Sequence 2924, Ap
c  15    27.6     21.4   2150  8  US-10-955-054A-98             Sequence 98, Appl
c  16    27.6     21.4   2150  9  US-11-072-175-35              Sequence 35, Appl
   17    27.4     21.2    398 12  US-11-000-688-1241            Sequence 1241, Ap
   18    27.4     21.2   1356  7  US-10-973-115B-43             Sequence 43, Appl
   19    27.4     21.2   1356  8  US-10-131-826A-43             Sequence 43, Appl
c  20    26.4     20.5    523  6  US-09-925-065A-108531         Sequence 108531,
   21    26.4     20.5   1223  6  US-09-925-065A-694721         Sequence 694721,
   22    26.4     20.5   1223  6  US-09-925-065A-694722         Sequence 694722,
   23    26.4     20.5   2499  8  US-10-750-185-28315           Sequence 28315, A
   24    26.4     20.5   2499  8  US-10-750-623-28315           Sequence 28315, A
c  25    26.2     20.3    495  6  US-09-925-065A-684535         Sequence 684535,
c  26    26.2     20.3    495  6  US-09-925-065A-684536         Sequence 684536,
c  27    26.2     20.3    578 12  US-11-128-061-1368            Sequence 1368, Ap
c  28    26.2     20.3    578 12  US-11-128-061-5010            Sequence 5010, Ap
c  29    26.2     20.3    578 12  US-11-128-049-1368            Sequence 1368, Ap
c  30    26.2     20.3    578 12  US-11-128-049-5010            Sequence 5010, Ap
c  31    26.2     20.3 121736  9  US-11-114-798-49              Sequence 49, Appl
c  32    26.2     20.3 318488  9  US-11-114-798-58              Sequence 58, Appl
   33      26     20.2    565  6  US-09-925-065A-663286         Sequence 663286,
   34      26     20.2    565  6  US-09-925-065A-663287         Sequence 663287,
c  35      26     20.2    600 12  US-11-136-527-7532            Sequence 7532, Ap
c  36      26     20.2   1025 12  US-11-136-527-3436            Sequence 3436, Ap
c  37      26     20.2   1426  6  US-09-925-065A-84896          Sequence 84896, A
c  38      26     20.2   1617  6  US-09-925-065A-81196          Sequence 81196, A
   39      26     20.2  33931 12  US-11-124-367A-5006           Sequence 5006, Ap
   40    25.8     20.0   1808  6  US-09-925-065A-679127         Sequence 679127,
c  41    25.6     19.8    463  6  US-09-925-065A-476394         Sequence 476394,
c  42    25.6     19.8    560  6  US-09-925-065A-251626         Sequence 251626,
   43    25.4     19.7    443  6  US-09-925-065A-239621         Sequence 239621,
   44    25.4     19.7    567  6  US-09-925-065A-770926         Sequence 770926,
c  45    25.4     19.7   2238  8  US-10-750-185-40094           Sequence 40094, A

                                ALIGNMENTS


RESULT 1
US-11-114-798-56
; Sequence 56, Application US/11114798
; Publication No. US20060035246A1
; GENERAL INFORMATION:
;  APPLICANT: WU, RINA
;  APPLICANT:  MARQUEZ, ABBEY
;  TITLE OF INVENTION: CHROMOGENIC IN SITU HYBRIDIZATION METHODS, KITS, AND
;  TITLE OF INVENTION:  COMPOSITIONS
;  FILE REFERENCE: 0618.011.0004
;  CURRENT APPLICATION NUMBER: US/11/114,798
;  CURRENT FILING DATE:  2005-04-26
;  PRIOR APPLICATION NUMBER: 10/173,525
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/952,851
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 56
;   LENGTH: 185750
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: nucleotide sequence
US-11-114-798-56

  Query Match             22.5%;  Score 29;  DB 9;  Length 185750;
  Best Local Similarity   55.4%;  Pred. No. 7.9;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
               |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db       70115 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 70174

Qy          76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                 |||    |   | | ||||||||  |   ||  || | |
Db       70175 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 70215


RESULT 2
US-10-928-446A-1
; Sequence 1, Application US/10928446A
```

```
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825234)
;   OTHER INFORMATION: the 'n' at position 825234 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825270)
;   OTHER INFORMATION: the 'n' at position 825270 may be 'c' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825401)
;   OTHER INFORMATION: the 'n' at position 825401 may be 'c' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825428)
;   OTHER INFORMATION: the 'n' at position 825428 may be 'g' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825473)
;   OTHER INFORMATION: the 'n' at position 825473 may be 'g' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825765)
;   OTHER INFORMATION: a "c" may be deleted at this position
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (825828)
;   OTHER INFORMATION: the 'n' at position 825828 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (826041)
;   OTHER INFORMATION: the 'n' at position 826041 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (826546)
;   OTHER INFORMATION: the 'n' at position 826546 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (826654)
;   OTHER INFORMATION: the 'n' at position 826654 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (826826)
;   OTHER INFORMATION: the 'n' at position 826826 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (826863)
;   OTHER INFORMATION: the 'n' at position 826863 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (827008)
;   OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (843055)
;   OTHER INFORMATION: the 'n' at position 843055 may be 'g' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (843118)
;   OTHER INFORMATION: the 'n' at position 843118 may be 't' or 'c'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (871027)
;   OTHER INFORMATION: the 'n' at position 871027 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (871140)
;   OTHER INFORMATION: the 'n' at position 871140 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (871168)
;   OTHER INFORMATION: the 'n' at position 871168 may be 'c' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (872678)
;   OTHER INFORMATION: the 'n' at position 872678 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (872742)
;   OTHER INFORMATION: the 'n' at position 872742 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (925859)
;   OTHER INFORMATION: the 'n' at position 925859 may be 'c' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (993220)
;   OTHER INFORMATION: the 'n' at position 993220 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (993254)
;   OTHER INFORMATION: the 'n' at position 993254 may be 'g' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1006462)
;   OTHER INFORMATION: the 'n' at position 1006462 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1007820)..(1007823)
;   OTHER INFORMATION: "ttct" may be deleted at this position
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1018038)
;   OTHER INFORMATION: the 'n' at position 1018038 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1018704)
;   OTHER INFORMATION: the 'n' at position 1018704 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1018718)..(1018720)
;   OTHER INFORMATION: "gtt" may be deleted at this position
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1026786)
;   OTHER INFORMATION: the 'n' at position 1026786 may be 'c' or 'a'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1047134)
;   OTHER INFORMATION: the 'n' at position 1047134 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1047159)
;   OTHER INFORMATION: the 'n' at position 1047159 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1047378)
;   OTHER INFORMATION: the 'n' at position 1047378 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
```

```
;   LOCATION: (1047739)
;   OTHER INFORMATION: the 'n' at position 1047739 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1050133)..(1050137)
;   OTHER INFORMATION: "ttaaa" may be deleted at this position
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1050539)
;   OTHER INFORMATION: the 'n' at position 1050539 may be 'c' or 't'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1062808)
;   OTHER INFORMATION: the 'n' at position 1062808 may be 'c' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1066392)
;   OTHER INFORMATION: the 'n' at position 1066392 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (1073711)
;   OTHER INFORMATION: the 'n' at position 1073711 may be 'c' or 't'
US-10-928-446A-1

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
               |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy          76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                 |||    |   | | ||||||||  |   ||  || | |
Db      771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 3
US-10-928-446A-181
; Sequence 181, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 181
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (826985)..(827008)
;   FEATURE:
;   OTHER INFORMATION: full exon 1 range is 826667-827008
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (827008)..(827008)
;   OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (843242)..(843315)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (922549)..(922630)
;   OTHER INFORMATION: exon
```

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (926021)..(926059)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (929123)..(929176)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (993104)..(993154)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999547)..(999608)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   OTHER INFORMATION: exon
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-181

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 4
US-10-928-446A-183
; Sequence 183, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (826985)..(827017)
;   FEATURE:
;   OTHER INFORMATION: full exon 1 range is 826667-827008
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (843242)..(843315)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (922549)..(922630)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (926021)..(926059)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (929123)..(929176)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (993104)..(993154)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999547)..(999608)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
```

```
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-183

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 5
US-10-928-446A-185
; Sequence 185, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
```

```
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 185
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (722487)..(722534)
;   FEATURE:
;   OTHER INFORMATION: full length exon 1 range is 722213-722534
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (843242)..(843315)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (922549)..(922630)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (926021)..(926059)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (929123)..(929176)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (993104)..(993154)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999547)..(999608)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-185

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 6
US-10-928-446A-187
```

```
; Sequence 187, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 187
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
```

```
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1044868)..(1044989)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1047519)..(1047589)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1050296)..(1050391)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1060368)..(1060441)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1062648)..(1062708)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1064561)..(1064620)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1066207)..(1066314)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1067768)..(1067864)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1068609)..(1068681)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1073289)..(1073388)
;    FEATURE:
;    OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-187

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 7
US-10-928-446A-189
; Sequence 189, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
;    LENGTH: 1080000
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (723508)..(723516)
;    FEATURE:
;    OTHER INFORMATION: full length exon 1 range is 723483-723516
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1000354)..(1000456)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1002118)..(1002284)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1006117)..(1006249)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1007860)..(1008036)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1010940)..(1011014)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1018160)..(1018291)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1018800)..(1018919)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1020028)..(1020225)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1026659)..(1026736)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1028113)..(1028167)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1034316)..(1034374)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1041390)..(1041455)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1043121)..(1043350)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1044868)..(1044989)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1047519)..(1047589)
;    OTHER INFORMATION: exon
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1050296)..(1050391)
;    OTHER INFORMATION: exon
;    FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-189

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 8
US-10-928-446A-191
; Sequence 191, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
```

```
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-191

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 9
US-10-928-446A-193
; Sequence 193, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 193
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   OTHER INFORMATION: exon
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
```

```
US-10-928-446A-193

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 10
US-10-928-446A-195
; Sequence 195, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 195
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-195

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 11
US-10-928-446A-197
; Sequence 197, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 197
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
```

```
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
```

```
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-197

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |    ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 12
US-10-928-446A-199
; Sequence 199, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 199
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   FEATURE:
;   OTHER INFORMATION: full length exon 7 range is 999547-999608
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
```

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
;   FEATURE:
;   OTHER INFORMATION: full length exon 30 range is 1073289-1075279
US-10-928-446A-199

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 13
US-10-928-446A-201
; Sequence 201, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
;  APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
;  TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
;  TITLE OF INVENTION:  VIRAL BUDDING
;  FILE REFERENCE: 0274-5785.1US
;  CURRENT APPLICATION NUMBER: US/10/928,446A
;  CURRENT FILING DATE:  2004-08-26
;  PRIOR APPLICATION NUMBER: 60/359,741
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 202
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 201
;   LENGTH: 1080000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: (827008)..(827008)
;   OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (872617)..(872860)
;   OTHER INFORMATION: Exon 2a
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (904952)..(905030)
;   OTHER INFORMATION: Exon 2d
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (922549)..(922630)
;   OTHER INFORMATION: Exon 3
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (926021)..(926059)
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (929123)..(929176)
;   OTHER INFORMATION: Exon 5
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (993104)..(993154)
;   OTHER INFORMATION: Exon 6
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (999562)..(999608)
;   OTHER INFORMATION: Full exon range is 999547-999608
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1000354)..(1000456)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1000354)..(1000456)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1002118)..(1002284)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1002118)..(1002284)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1006117)..(1006249)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1006117)..(1006249)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1007860)..(1008036)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1007860)..(1008036)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1010940)..(1011014)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1010940)..(1011014)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1018160)..(1018291)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018160)..(1018291)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1018800)..(1018919)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1018800)..(1018919)
;   FEATURE:
```

```
;   NAME/KEY: exon
;   LOCATION: (1020028)..(1020225)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1020028)..(1020225)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1026659)..(1026736)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1026659)..(1026736)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1028113)..(1028167)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1028113)..(1028167)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1034316)..(1034374)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1034316)..(1034374)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1041390)..(1041455)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1041390)..(1041455)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1043121)..(1043350)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1043121)..(1043350)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1044868)..(1044989)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1044868)..(1044989)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1047519)..(1047589)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1047519)..(1047589)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1050296)..(1050391)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1050296)..(1050391)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1060368)..(1060441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1060368)..(1060441)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1062648)..(1062708)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1062648)..(1062708)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1064561)..(1064620)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1064561)..(1064620)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1066207)..(1066314)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1066207)..(1066314)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1067768)..(1067864)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1067768)..(1067864)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1068609)..(1068681)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1068609)..(1068681)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1073289)..(1075279)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1073289)..(1073388)
US-10-928-446A-201

  Query Match             22.5%;  Score 29;  DB 8;  Length 1080000;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db     771426 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 771485

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db     771486 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 771526


RESULT 14
US-11-124-368A-2924/c
; Sequence 2924, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
;  APPLICANT: Michele Cargill
;  APPLICANT:  James J. Devlin
;  APPLICANT:  May Luke
;  TITLE OF INVENTION: Genetic Polymorphisms Associated with
;  TITLE OF INVENTION:  Vascular Diseases, Methods of Detection and Uses Thereof
;  FILE REFERENCE: CL001524
;  CURRENT APPLICATION NUMBER: US/11/124,368A
;  CURRENT FILING DATE:  2005-05-09
;  PRIOR APPLICATION NUMBER: US 60/568,845
;  PRIOR FILING DATE: 2004-05-07
;  PRIOR APPLICATION NUMBER: US 60/625,936
;  PRIOR FILING DATE: 2004-11-09
;  NUMBER OF SEQ ID NOS: 21112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2924
;   LENGTH: 79134
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-124-368A-2924

  Query Match             22.3%;  Score 28.8;  DB 12;  Length 79134;
  Best Local Similarity   58.0%;  Pred. No. 7.7;
  Matches   51;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy         23 ACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTG 82
              | ||| || ||||| |||   | | ||| || |||   || |      |||||| || |
Db      76157 AGTTCTAGGATTACAGACATGAGACACCACGCCCGGCCTGATATATAACTTTTTTTTTTT 76098

Qy         83 GTGCTGTGACGTGTTGTCCAACCGTATT 110
               |  | |||  ||  |||  ||  | ||
Db      76097 TTTTTTTGAGATGAAGTCTCACTCTGTT 76070
```

```
RESULT 15
US-10-955-054A-98/c
; Sequence 98, Application US/10955054A
; Publication No. US20050266420A1
; GENERAL INFORMATION:
;  APPLICANT: PUSZTAI, LAJOS
;  APPLICANT:  SYMMANS, W. FRASER
;  APPLICANT:  HESS, KENNETH R.
;  APPLICANT:  AYERS, MARK
;  APPLICANT:  STEC, JAMES
;  TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
;  FILE REFERENCE: UTXC:880US
;  CURRENT APPLICATION NUMBER: US/10/955,054A
;  CURRENT FILING DATE:  2004-09-30
;  NUMBER OF SEQ ID NOS: 195
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
;   LENGTH: 2150
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-955-054A-98

  Query Match             21.4%;  Score 27.6;  DB 8;  Length 2150;
  Best Local Similarity   52.6%;  Pred. No. 8.8;
  Matches   60;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          7 TCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCG 66
              ||||||   |   ||| |||   |  ||| |      ||| | | | || |||    ||
Db       1239 TCTGCTGCAATTCATTTCTTTCATGATTTATTCGCTCAATGTCTGCAACTGAGTACTTCT 1180

Qy         67 GGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              || | |   || | |  ||  ||   ||||| |||     |||| || |  | |
Db       1179 GGTTGTCAATGATATTCTGTAGTCGAGTGTTCTCCTGTTTTATTGTTTCACATT 1126


Search completed: March  7, 2006, 13:20:56
Job time : 1206.75 secs
```

```
                    GenCore version 5.1.7
           Copyright (c) 1993 - 2006  Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:         March  7, 2006, 12:14:00 ; Search time 354.047 Seconds
                                           (without alignments)
                                           3013.020 Million cell updates/sec

Title:          US-10-782-899-1_COPY_472_600
Perfect score:  129
Sequence:       1 tgaccttctgctcgtagcga..........tattccggactagttcagcg 129

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters:      19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA_Main:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result           Query
  No.    Score   Match Length DB  ID                          Description
 ------------------------------------------------------------------------------
     1     129   100.0    600  3  US-09-816-391A-1             Sequence 1, Appli
     2     129   100.0    600  8  US-10-782-899-1              Sequence 1, Appli
c    3     129   100.0 2256646  7  US-10-470-565-1              Sequence 1, Appli
     4    29.4    22.8    885  9  US-10-617-320-306            Sequence 306, App
c    5    29.4    22.8   1377  7  US-10-687-677-1              Sequence 1, Appli
     6    29.4    22.8   3520  3  US-09-814-353-19115          Sequence 19115, A
c    7    29.2    22.6    634  7  US-10-332-859-15             Sequence 15, Appl
     8    29.2    22.6   1013  7  US-10-332-859-232            Sequence 232, App
     9    29.2    22.6   1836  7  US-10-332-859-17             Sequence 17, Appl
    10      29    22.5 344805  8  US-10-779-271-1              Sequence 1, Appli
    11      29    22.5 354592  9  US-10-737-082-70             Sequence 70, Appl
    12      29    22.5 354592  9  US-10-765-790-70             Sequence 70, Appl
    13    28.8    22.3    766  7  US-10-332-859-317            Sequence 317, App
c   14    28.6    22.2   1043  8  US-10-425-115-152385         Sequence 152385,
    15    28.4    22.0    936  3  US-09-903-410-31             Sequence 31, Appl
    16    28.4    22.0    936  5  US-10-027-805-31             Sequence 31, Appl
    17    28.4    22.0    936  5  US-10-027-804-31             Sequence 31, Appl
    18    28.4    22.0   2451  3  US-09-938-842A-64            Sequence 64, Appl
    19    28.4    22.0   2451  3  US-09-938-842A-64            Sequence 64, Appl
c   20    28.2    21.9    599  6  US-10-029-386-13097          Sequence 13097, A
    21    28.2    21.9   1152  9  US-10-450-763-9386           Sequence 9386, Ap
c   22    28.2    21.9   3519  7  US-10-437-963-80607          Sequence 80607, A
c   23      28    21.7 165221  5  US-10-087-192-1015           Sequence 1015, Ap
```

```
c   24    27.8    21.6    483  7  US-10-424-599-128897         Sequence 128897,
c   25    27.8    21.6   2037  5  US-10-052-586-591            Sequence 591, App
c   26    27.8    21.6   2037  5  US-10-174-590-591            Sequence 591, App
c   27    27.8    21.6   2037  5  US-10-176-758-591            Sequence 591, App
c   28    27.8    21.6   2037  5  US-10-175-737-591            Sequence 591, App
c   29    27.8    21.6   2037  5  US-10-174-581-591            Sequence 591, App
c   30    27.8    21.6   2037  5  US-10-176-483-591            Sequence 591, App
c   31    27.8    21.6   2037  5  US-10-176-749-591            Sequence 591, App
c   32    27.8    21.6   2037  5  US-10-176-914-591            Sequence 591, App
c   33    27.8    21.6   2037  5  US-10-176-915-591            Sequence 591, App
c   34    27.8    21.6   2037  5  US-10-173-706-591            Sequence 591, App
c   35    27.8    21.6   2037  5  US-10-175-738-591            Sequence 591, App
c   36    27.8    21.6   2037  5  US-10-175-752-591            Sequence 591, App
c   37    27.8    21.6   2037  5  US-10-176-482-591            Sequence 591, App
c   38    27.8    21.6   2037  5  US-10-176-757-591            Sequence 591, App
c   39    27.8    21.6   2037  5  US-10-176-913-591            Sequence 591, App
c   40    27.8    21.6   2037  5  US-10-180-552-591            Sequence 591, App
c   41    27.8    21.6   2037  5  US-10-180-557-591            Sequence 591, App
c   42    27.8    21.6   2037  5  US-10-173-700-591            Sequence 591, App
c   43    27.8    21.6   2037  5  US-10-174-572-591            Sequence 591, App
c   44    27.8    21.6   2037  5  US-10-174-579-591            Sequence 591, App
c   45    27.8    21.6   2037  5  US-10-174-582-591            Sequence 591, App

                              ALIGNMENTS


RESULT 1
US-09-816-391A-1
; Sequence 1, Application US/09816391A
; Patent No. US20020054865A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, Minoru
;  APPLICANT:  TANIGUCHI, Shunichiro
;  APPLICANT:  AMANO, Jun
;  APPLICANT:  YAZAWA, Kazuyuki
;  APPLICANT:  KANO, Yasunobu
;  APPLICANT:  NAKAMURA, Toshiyuki
;  APPLICANT:  SASAKI, Takayuki
;  TITLE OF INVENTION: Anaerobic bacterium as a drug for cancer gene therapy
;  FILE REFERENCE: 2001-WMC/01736
;  CURRENT APPLICATION NUMBER: US/09/816,391A
;  CURRENT FILING DATE:  2001-03-26
;  PRIOR APPLICATION NUMBER: JP 00/287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-09-816-391A-1

  Query Match              100.0%;  Score 129;  DB 3;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 5e-37;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600


RESULT 2
```

```
US-10-782-899-1
; Sequence 1, Application US/10782899
; Publication No. US20050025745A1
; GENERAL INFORMATION:
;  APPLICANT: FUJIMORI, MINORU
;  APPLICANT:  TANIGUCHI, SHUNICHIRO
;  APPLICANT:  AMANO, JUN
;  APPLICANT:  YAZAWA, KAZUYUKI
;  APPLICANT:  KANO, YASUNOBU
;  APPLICANT:  NAKAMURA, TOSHIYUKI
;  APPLICANT:  SASAKI, TAKAYUKI
;  TITLE OF INVENTION: ANAEROBIC BACTERIUM AS A DRUG FOR CANCER GENE THERAPY
;  FILE REFERENCE: 671308-2001.1
;  CURRENT APPLICATION NUMBER: US/10/782,899
;  CURRENT FILING DATE:  2004-02-23
;  PRIOR APPLICATION NUMBER: 09/816,391
;  PRIOR FILING DATE: 2001-03-26
;  PRIOR APPLICATION NUMBER: JP 2000-287688
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (193)..(471)
US-10-782-899-1

  Query Match             100.0%;  Score 129;  DB 8;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 5e-37;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 531

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 591

Qy        121 AGTTCAGCG 129
              |||||||||
Db        592 AGTTCAGCG 600

RESULT 3
US-10-470-565-1/c
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
;  APPLICANT: Societe des Produits Nestle S.A.
;  TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
;  FILE REFERENCE: 80290/WO
;  CURRENT APPLICATION NUMBER: US/10/470,565
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: EP 01102050.0
;  PRIOR FILING DATE: 2001-01-30
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 2256646
;   TYPE: DNA
;   ORGANISM: Bifidobacterium longum
US-10-470-565-1

  Query Match             100.0%;  Score 129;  DB 7;  Length 2256646;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-36;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    2241699 TGACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGA 2241640

Qy         61 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2241639 TGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGGACT 2241580

Qy        121 AGTTCAGCG 129
              |||||||||
Db    2241579 AGTTCAGCG 2241571


RESULT 4
US-10-617-320-306
; Sequence 306, Application US/10617320
; Publication No. US20050136404A1
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGN
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/617,320
;             FILING DATE: 10-Jul-2003
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 306:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 885 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...885
;        SEQUENCE DESCRIPTION: SEQ ID NO: 306:
US-10-617-320-306

  Query Match             22.8%;  Score 29.4;  DB 9;  Length 885;
  Best Local Similarity   58.6%;  Pred. No. 3.5;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         32 ATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGA 91
              ||||| ||   |||  |  |  ||| | || || | |  |||| ||||| | || || |
```

```
Db        604 ATTACAGATATCAACAAGGCTTACCTAAATCGTGGTGACCTTTCTGTTGAGCTGATGGGG 663

Qy         92 CGTGTTGTCCAACCGTATTATTCCGGA 118
              |||| | |      ||   || | |||
Db        664 CGTGGTTTTGCCTGGTTGGATACGGGA 690

RESULT 5
US-10-687-677-1/c
; Sequence 1, Application US/10687677
; Publication No. US20040142419A1
; GENERAL INFORMATION:
;  APPLICANT: Guy, John
;  TITLE OF INVENTION: REDUCING CELLULAR DYSFUNCTION CAUSED BY MITOCHONDRIAL GENE
;  TITLE OF INVENTION:  MUTATIONS
;  FILE REFERENCE: 5853-324
;  CURRENT APPLICATION NUMBER: US/10/687,677
;  CURRENT FILING DATE:  2003-10-17
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1377
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-687-677-1

  Query Match             22.8%;  Score 29.4;  DB 7;  Length 1377;
  Best Local Similarity   63.4%;  Pred. No. 4;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         57 GAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
              | || | | ||  || | ||||||  ||| |||| |  | || || || || || ||  |
Db       1292 GTGAAGCTGGGCTTCATGTTGTTGATGTGGTGTGTCAGGCTGCCCCACTGTGTTGTTGTG 1233

Qy        117 GACTAGTTCAG 127
               ||  || |||
Db       1232 AACATGTACAG 1222

RESULT 6
US-09-814-353-19115
; Sequence 19115, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19115
;   LENGTH: 3520
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-19115
```

```
  Query Match             22.8%;  Score 29.4;  DB 3;  Length 3520;
  Best Local Similarity   54.1%;  Pred. No. 5.4;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy          4 CCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGG 63
              || ||| ||  |      | |||      ||  | |    |   | |||   |  | |||
Db         27 CCCTCTACTGCTCATCTCTGCTTACCTTGTTTTTTAGTGGATTAATCCCTCACCTGGTGG 86

Qy         64 TCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTC 114
               ||  ||  |||| |||||| ||||| || || || || ||   || || |
Db         87 GCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTCATAATCC 137

RESULT 7
US-10-332-859-15/c
; Sequence 15, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
;   LENGTH: 634
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-15

  Query Match             22.6%;  Score 29.2;  DB 7;  Length 634;
  Best Local Similarity   59.8%;  Pred. No. 3.8;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        245 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 186

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        185 ATCAATGCCTACTTGGTGGACG 164

RESULT 8
US-10-332-859-232
; Sequence 232, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
```

```
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 232
;   LENGTH: 1013
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-232

  Query Match             22.6%;  Score 29.2;  DB 7;  Length 1013;
  Best Local Similarity   59.8%;  Pred. No. 4.3;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        769 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 828

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        829 ATCAATGCCTACTTGGTGGACG 850

RESULT 9
US-10-332-859-17
; Sequence 17, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
;   LENGTH: 1836
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-17

  Query Match             22.6%;  Score 29.2;  DB 7;  Length 1836;
  Best Local Similarity   59.8%;  Pred. No. 5.2;
  Matches   49;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
              | |||   |    ||| ||  ||  |||| |||||| ||||| || || || || ||
Db        813 ATCCCTCACCTCCTGGGCGATGTGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 872

Qy        108 ATTATTCCGGACTAGTTCAGCG 129
              || | | |  ||| | |   ||
Db        873 ATCAATGCCTACTTGGTGGACG 894

RESULT 10
US-10-779-271-1
; Sequence 1, Application US/10779271
; Publication No. US20040220387A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: AstraZeneca AB
;  TITLE OF INVENTION: Methods
;  FILE REFERENCE: ASZD-P02-251
;  CURRENT APPLICATION NUMBER: US/10/779,271
;  CURRENT FILING DATE:  2004-02-13
;  PRIOR APPLICATION NUMBER: 09/463,844
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: PCT/GB98/00259
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 9716162.4
;  PRIOR FILING DATE: 1997-08-97
;  PRIOR APPLICATION NUMBER: 60/535,986
;  PRIOR FILING DATE: 2004-01-12
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 344805
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-779-271-1

  Query Match             22.5%;  Score 29;  DB 8;  Length 344805;
  Best Local Similarity   55.4%;  Pred. No. 32;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      47662 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 47721

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db      47722 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 47762

RESULT 11
US-10-737-082-70
; Sequence 70, Application US/10737082
; Publication No. US20050130170A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2032
;  CURRENT APPLICATION NUMBER: US/10/737,082
;  CURRENT FILING DATE:  2003-12-16
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
;   LENGTH: 354592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-737-082-70

  Query Match             22.5%;  Score 29;  DB 9;  Length 354592;
  Best Local Similarity   55.4%;  Pred. No. 32;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy         16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
              |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db      50053 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 50112

Qy         76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                |||    |   | | ||||||||  |   ||  || | |
Db      50113 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 50153
```

```
RESULT 12
US-10-765-790-70
; Sequence 70, Application US/10765790
; Publication No. US20050130172A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer Healthcare LLC
;  APPLICANT:  Beard, Chris
;  APPLICANT:  Burgess, Chris
;  APPLICANT:  Gannon, Allison
;  APPLICANT:  Harvey, Jeanne
;  APPLICANT:  Lechner, John F.
;  APPLICANT:  Li, Zheng
;  TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;  FILE REFERENCE: 1657/2035
;  CURRENT APPLICATION NUMBER: US/10/765,790
;  CURRENT FILING DATE:  2004-01-27
;  PRIOR APPLICATION NUMBER: US 10/737,082
;  PRIOR FILING DATE: 2003-12-16
;  NUMBER OF SEQ ID NOS: 300
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
;   LENGTH: 354592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-765-790-70

  Query Match             22.5%;  Score 29;  DB 9;  Length 354592;
  Best Local Similarity   55.4%;  Pred. No. 32;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy          16 AGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTT 75
               |||  ||   | |||  || || |||  ||  |   |||   || |||| |  ||||| |
Db       50053 AGCAGTTCACTTGAGACTTTCTCACGGAAAGAAGTGCGATACAGTTGGTGGTAGTCTTAT 50112

Qy          76 TGTTGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCG 116
                 |||    |   | | ||||||||  |   ||  || | |
Db       50113 ATTTGCTACGTGCTCATGTGTTGTCTGATGATAGAATGCAG 50153

RESULT 13
US-10-332-859-317
; Sequence 317, Application US/10332859
; Publication No. US20040088746A1
; GENERAL INFORMATION:
;  APPLICANT: Grimm, Stefan
;  APPLICANT:  Schoenfeld, Nicole
;  APPLICANT:  Braziulis, Erik
;  APPLICANT:  Cramer, Ursula
;  APPLICANT:  Gewies, Andreas
;  APPLICANT:  Voss, Frank
;  APPLICANT:  Mund, Thomas
;  APPLICANT:  Albayrak, Timur
;  APPLICANT:  Gille, Hendrik
;  APPLICANT:  Klein, Matthias
;  APPLICANT:  Bauer, Manuel
;  TITLE OF INVENTION: Apoptosis-Inducing DNA Sequences
;  FILE REFERENCE: 2923-0133
;  CURRENT APPLICATION NUMBER: US/10/332,859
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: PCT/EP01/08170
;  PRIOR FILING DATE: 2001-07-13
;  NUMBER OF SEQ ID NOS: 355
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 317
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-332-859-317

  Query Match             22.3%;  Score 28.8;  DB 7;  Length 766;
  Best Local Similarity   58.5%;  Pred. No. 5.6;
  Matches   48;  Conservative    1;  Mismatches   33;  Indels    0;  Gaps    0;
```

```
Qy          48 ACCCCGACCGAGATGGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTGTTGTCCAACCGT 107
               | |||   |    ||| ||  |:  |||| |||||| ||||| || || || || ||
Db         229 ATCCCTCACCTCCTGGGCGATGWGGTTTTCTTGTGGGGCTGTAACCTGCTGGCCCACTTC 288

Qy         108 ATTATTCCGGACTAGTTCAGCG 129
               || | | |  ||| | |   ||
Db         289 ATCAATGCCTACTTGGTGGACG 310

RESULT 14
US-10-425-115-152385/c
; Sequence 152385, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 152385
;   LENGTH: 1043
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_70558C.1
US-10-425-115-152385

  Query Match             22.2%;  Score 28.6;  DB 8;  Length 1043;
  Best Local Similarity   55.6%;  Pred. No. 7.3;
  Matches   55;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy          19 GATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGATGGTCGGGGTCTTTTTGT 78
               | | |||||   | ||  ||||   |  || | ||||  ||  ||||||| |||   ||
Db        1041 GCTGACTTCCCACCTTGATGACTCAAGGGAGCACGACGTAGTAGGTCGGGATCTGGGTGC 982

Qy          79 TGTGGTGCTGTGACGTGTTGTCCAACCGTATTATTCCGG 117
               |   | ||   | ||| ||  | || |      |||| |
Db         981 TCCTGAGCATCGGCGTCTTCACGAAACCCGAACTTCCCG 943

RESULT 15
US-09-903-410-31
; Sequence 31, Application US/09903410
; Patent No. US20020146799A1
; GENERAL INFORMATION:
;  APPLICANT: DIVERSA CORPORATION
;  APPLICANT:  ROBERTSON, Dan
;  APPLICANT:  MURPHY, Dennis
;  APPLICANT:  REID, John
;  APPLICANT:  MAFFIA, Anthony
;  APPLICANT:  LINK, Steven
;  APPLICANT:  SWANSON, Ronald
;  APPLICANT:  WARREN, Patrick
;  APPLICANT:  KOSMOTKA, Anna
;  TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
;  FILE REFERENCE: DIVER1180-2
;  CURRENT APPLICATION NUMBER: US/09/903,410
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: US 09/382,242
;  PRIOR FILING DATE: 1999-08-24
;  PRIOR APPLICATION NUMBER: US 08/602,359
;  PRIOR FILING DATE: 1996-02-16
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 936
;   TYPE: DNA
```

```
;   ORGANISM: Archaeoglobus fulgidus
US-09-903-410-31

  Query Match            22.0%;  Score 28.4;  DB 3;  Length 936;
  Best Local Similarity  56.4%;  Pred. No. 8.4;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          2 GACCTTCTGCTCGTAGCGATTACTTCGAGCATTACTGACGACAAAGACCCCGACCGAGAT 61
              ||||||  |  | || |   | |   || ||| || | |||  | |||||     |||||
Db        718 GACCTTGAGAACCTACCTCCTGCGCTGATCATAACCGCCGAATACGACCCGCTGAGAGAT 777

Qy         62 GGTCGGGGTCTTTTTGTTGTGGTGCTGTGACGTG 95
              |   | |   |||| |    | ||||| || | |
Db        778 GAAGGAGAAGTTTTCGGGCAGATGCTGAGAAGAG 811


Search completed: March  7, 2006, 13:34:34
Job time : 358.047 secs
```